

UNIVERSITY  
LIBRARY  
NOTTINGHAM

28/6

THE LIBRARY	
UNIVERSITY OF NOTTINGHAM	
Class Mark	s/oc 6 <del>37</del>
Book Number	128674

SEP 1965

28/6

UNIVERSITY OF NOTTINGHAM  
**WITHDRAWN**  
FROM THE LIBRARY



61 0105472 4



UNIVERSITY OF BIRMINGHAM  
WITHDRAWN  
FROM THE LIBRARY

Students and External Readers

Staff & Research  
Students

DATE DUE FOR RETURN

DATE OF ISSUE

~~12 OCT 1982~~

~~12 OCT 1982~~

N.B. All books  
must be returned  
for the Annual  
Inspection in  
June

Any book which you borrow remains your responsibility  
until the loan slip is cancelled





**THE LOGIC OF MODERN PHYSICS**



THE MACMILLAN COMPANY  
NEW YORK • BOSTON • CHICAGO • DALLAS  
ATLANTA • SAN FRANCISCO

MACMILLAN AND CO., LIMITED  
LONDON • BOMBAY • CALCUTTA • MADRAS  
MELBOURNE

THE MACMILLAN COMPANY  
OF CANADA, LIMITED  
TORONTO

# THE LOGIC OF MODERN PHYSICS

BY

P. W. BRIDGMAN

HOLLIS PROFESSOR OF MATHEMATICS AND NATURAL HISTORY  
IN HARVARD UNIVERSITY



New York

THE MACMILLAN COMPANY



COPYRIGHT, 1927,  
By THE MACMILLAN COMPANY.

---

All rights reserved—no part of this book may be reproduced in any form without permission in writing from the publisher, except by a reviewer who wishes to quote brief passages in connection with a review written for inclusion in magazine or newspaper.

---

Set up and electrotyped.  
Published April, 1927.  
Sixth Printing, 1951

PRINTED IN THE UNITED STATES OF AMERICA

## PREFACE

This excursion into the field of fundamental criticism by one whose activities have hitherto been confined almost entirely to experiment is not evidence of senile decay, as might be cynically assumed. I have always, throughout all my experimental work, felt an imperative need of a better understanding of the foundations of our physical thought and have for a long time made more or less unsystematic attempts to reach such an understanding. Only now, however, has a half sabbatical year given me leisure to attempt a more or less orderly exposition.

In spite of previous writings on the broad fundamentals by Clifford, Stallo, Mach, and Poincaré, to mention only a few, I believe a new essay of this critical character needs no apology. For entirely apart from the question of whether many of the points of view of these essays can be maintained, the discovery of new facts in the domain of relativity and quantum theory has shifted the center of interest and emphasis. All the quite recent activity with the new quantum mechanics seems to call for a new examination of fundamental matters which shall recognize, at least by implication, the existence of the special phenomena of the quantum domain. However, the necessity for re-examination does not mean at all that many of

the results of previous criticism may not still be accepted; some of these results have become so thoroughly incorporated into physical thinking that we can assume them without mention. Thus the fundamental attitude of this essay is empiricism, which is now justified as the attitude of the physicist in large part by the inquiry into the physiological origin of our concepts of space, time, and mechanics with which the previous essays were largely concerned.

None of the previous essays have consciously or immediately affected the details of this; in fact I have not read any of them within several years. If passages here recall passages already written, it is because the ideas have been assimilated and the precise origin forgotten; it is probably worth while to let such passages stand without revision, because such ideas gain in plausibility through having been found acceptable to independent thought.

I am much indebted to Professor R. F. Alfred Hoernlé of the Department of Philosophy of Johannesburg University, South Africa, for suggesting several modifications to make the text more acceptable to a philosopher, and slight amplifications for the benefit of readers not familiar with all the details of recent technical developments in physics.



## INTRODUCTION

One of the most noteworthy movements in recent physics is a change of attitude toward what may be called the interpretative aspect of physics. It is being increasingly recognized, both in the writings and the conversation of physicists, that the world of experiment is not understandable without some examination of the purpose of physics and of the nature of its fundamental concepts. It is no new thing to attempt a more critical understanding of the nature of physics, but until recently all such attempts have been regarded with a certain suspicion or even sometimes contempt. The average physicist is likely to deprecate his own concern with such questions, and is inclined to dismiss the speculations of fellow physicists with the epithet "metaphysical." This attitude has no doubt had a certain justification in the utter unintelligibility to the physicist of many metaphysical speculations and the sterility of such speculations in yielding physical results. However, the growing reaction favoring a better understanding of the interpretative fundamentals of physics is not a pendulum swing of the fashion of thought toward metaphysics, originating in the upheaval of moral values produced by the great war, or anything of the sort, but is a

reaction absolutely forced upon us by a rapidly increasing array of cold experimental facts.

This reaction, or rather new movement, was without doubt initiated by the restricted theory of relativity of Einstein. Before Einstein, an ever increasing number of experimental facts concerning bodies in rapid motion required increasingly complicated modifications in our naïve notions in order to preserve self-consistency, until Einstein showed that everything could be restored again to a wonderful simplicity by a slight change in some of our fundamental concepts. The concepts which were most obviously touched by Einstein were those of space and time, and much of the writing consciously inspired by Einstein has been concerned with these concepts. But that experiment compels a critique of much more than the concepts of space and time is made increasingly evident by all the new facts being discovered in the quantum realm.

The situation presented to us by these new quantum facts is two-fold. In the first place, all these experiments are concerned with things so small as to be forever beyond the possibility of direct experience, so that we have the problem of translating the evidence of experiment into other language. Thus we observe an emission line in a spectroscope and may infer an electron jumping from one energy level to another in an atom. In the second place, we have the problem of understanding the translated experimental evidence. Now of course every one knows that this problem is making us the greatest difficulty.

The experimental facts are so utterly different from those of our ordinary experience that not only do we apparently have to give up generalizations from past experience as broad as the field equations of electrodynamics, for instance, but it is even being questioned whether our ordinary forms of thought are applicable in the new domain; it is often suggested, for example, that the concepts of space and time break down.

The situation is rapidly becoming acute. Since I began writing this essay, there has been a striking increase in critical activity inspired by the new quantum mechanics of 1925-26, and it is common to hear expositions of the new ideas prefaced by analysis of what experiment really gives to us or what our fundamental concepts really mean. The change in ideas is now so rapid that a number of the statements of this essay are already antiquated as expressions of the best current opinion; however I have allowed these statements to stand, since the fundamental arguments are in nowise affected and we have no reason to think that present best opinions are in any way final. We have the impression of being in an important formative period; if we are, the complexion of physics for a long time in the future will be determined by our present attitude toward fundamental questions of interpretation. To meet this situation it seems to me that something more is needed than the hand-to-mouth philosophy that is now growing up to meet special emergencies, something approaching more closely to a systematic philosophy of all physics



which shall cover the experimental domains already consolidated as well as those which are now making us so much trouble. It is the attempt of this essay to give a more or less inclusive critique of all physics. Our problem is the double one of understanding what we are trying to do and what our ideals should be in physics, and of understanding the nature of the structure of physics as it now exists. These two ends are together furthered by an analysis of the fundamental concepts of physics; an understanding of the concepts we now have discloses the present structure of physics and a realization of what the concepts should be involves the ideals of physics. This essay will be largely concerned with the fundamental concepts; it will appear that almost all the concepts can profit from re-examination.

The material of this essay is largely obtained by observation of the actual currents of opinion in physics; much of what I have to say is more or less common property and doubtless every reader will find passages that he will feel have been taken out of his own mouth. On certain broad tendencies in present day physics, however, I have put my own interpretation, and it is more than likely that this interpretation will be unacceptable to many. But even if not acceptable, I hope that the stimulus of combatting the ideas offered here may be of value.

Certain limitations will have to be set to our inquiry in order to keep it within manageable compass. It is of course the merest truism that all our experimental knowledge and our understanding of

nature is impossible and non-existent apart from our own mental processes, so that strictly speaking no aspect of psychology or epistemology is without pertinence. Fortunately we shall be able to get along with a more or less naïve attitude toward many of these matters. We shall accept as significant our common sense judgment that there is a world external to us, and shall limit as far as possible our inquiry to the behavior and interpretation of this "external" world. We shall rule out inquiries into our states of consciousness as such. In spite, however, of the best intentions, we shall not be able to eliminate completely considerations savoring of the metaphysical, because it is evident that the nature of our thinking mechanism essentially colors any picture that we can form of nature, and we shall have to recognize that unavoidable characteristics of any outlook of ours are imposed in this way.





# CONTENTS

	PAGE
PREFACE . . . . .	v
INTRODUCTION . . . . .	vii
CHAPTER	
I. BROAD POINTS OF VIEW . . . . .	I
New Kinds of Experience Always Possible . .	2
The Operational Character of Concepts . . .	3
Einstein's Contribution in Changing Our Atti- tude toward Concepts . . . . .	3
Detailed Discussion of the Concept of Length .	9
The Relative Character of Knowledge . . .	25
Meaningless Questions . . . . .	28
General Comments on the Operational Point of View . . . . .	31
II. OTHER GENERAL CONSIDERATIONS . . . . .	33
The Approximate Character of Empirical Knowl- edge . . . . .	33
Explanations and Mechanisms . . . . .	37
Models and Constructs . . . . .	52
The Rôle of Mathematics in Physics . . . . .	60
III. DETAILED CONSIDERATION OF VARIOUS CONCEPTS OF PHYSICS . . . . .	66
The Concept of Space . . . . .	66
The Concept of Time . . . . .	68
The Causality Concept . . . . .	80
The Concept of Identity . . . . .	91
The Concept of Velocity . . . . .	97
The Concepts of Force and Mass . . . . .	102
The Concept of Energy . . . . .	108
The Concepts of Thermodynamics . . . . .	117
Electrical Concepts . . . . .	131

	PAGE
The Nature of Light and the Concepts of Relativity	150
Other Relativity Concepts . . . . .	166
Rotational Motion and Relativity . . . . .	178
Quantum Concepts . . . . .	186
 CHAPTER	
IV. SPECIAL VIEWS OF NATURE . . . . .	196
The Simplicity of Nature . . . . .	198
Determinism . . . . .	211
On the Possibility of Describing Nature Completely in Terms of Analysis . . . . .	220
A Glimpse Ahead . . . . .	222
INDEX . . . . .	227

# THE LOGIC OF MODERN PHYSICS



# THE LOGIC OF MODERN PHYSICS

## CHAPTER I

### BROAD POINTS OF VIEW

WHATEVER may be one's opinion as to our permanent acceptance of the analytical details of Einstein's restricted and general theories of relativity, there can be no doubt that through these theories physics is permanently changed. It was a great shock to discover that classical concepts, accepted unquestioningly, were inadequate to meet the actual situation, and the shock of this discovery has resulted in a critical attitude toward our whole conceptual structure which must at least in part be permanent. Reflection on the situation after the event shows that it should not have needed the new experimental facts which led to relativity to convince us of the inadequacy of our previous concepts, but that a sufficiently shrewd analysis should have prepared us for at least the possibility of what Einstein did.

Looking now to the future, our ideas of what external nature is will always be subject to change as we gain new experimental knowledge, but there is a part of our attitude to nature which should not be subject to future change, namely that part which rests on the permanent basis of the character of our



minds. It is precisely here, in an improved understanding of our mental relations to nature, that the permanent contribution of relativity is to be found. We should now make it our business to understand so thoroughly the character of our permanent mental relations to nature that another change in our attitude, such as that due to Einstein, shall be forever impossible. It was perhaps excusable that a revolution in mental attitude should occur once, because after all physics is a young science, and physicists have been very busy, but it would certainly be a reproach if such a revolution should ever prove necessary again.

#### NEW KINDS OF EXPERIENCE ALWAYS POSSIBLE

The first lesson of our recent experience with relativity is merely an intensification and emphasis of the lesson which all past experience has also taught, namely, that when experiment is pushed into new domains, we must be prepared for new facts, of an entirely different character from those of our former experience. This is taught not only by the discovery of those unsuspected properties of matter moving with high velocities, which inspired the theory of relativity, but also even more emphatically by the new facts in the quantum domain. To a certain extent, of course, the recognition of all this does not involve a change of former attitude; the *fact* has always been for the physicist the one ultimate thing from which there is no appeal, and in the face of which the only possible attitude is a humility almost

religious. The new feature in the present situation is an intensified conviction that in reality new orders of experience do exist, and that we may expect to meet them continually. We have already encountered new phenomena in going to high velocities, and in going to small scales of magnitude: we may similarly expect to find them, for example, in dealing with relations of cosmic magnitudes, or in dealing with the properties of matter of enormous densities, such as is supposed to exist in the stars.

Implied in this recognition of the possibility of new experience beyond our present range, is the recognition that no element of a physical situation, no matter how apparently irrelevant or trivial, may be dismissed as without effect on the final result until proved to be without effect by actual experiment.

The attitude of the physicist must therefore be one of pure empiricism. He recognizes no *a priori* principles which determine or limit the possibilities of new experience. Experience is determined only by experience. This practically means that we must give up the demand that all nature be embraced in any formula, either simple or complicated. It may perhaps turn out eventually that as a matter of fact nature can be embraced in a formula, but we must so organize our thinking as not to demand it as a necessity.

#### THE OPERATIONAL CHARACTER OF CONCEPTS

#### *Einstein's Contribution in Changing Our Attitude Toward Concepts*

Recognizing the essential unpredictability of

experiment beyond our present range, the physicist, if he is to escape continually revising his attitude, must use in describing and correlating nature concepts of such a character that our present experience does not exact hostages of the future. Now here it seems to me is the greatest contribution of Einstein. Although he himself does not explicitly state or emphasize it, I believe that a study of what he has done will show that he has essentially modified our view of what the concepts useful in physics are and should be. Hitherto many of the concepts of physics have been defined in terms of their properties. An excellent example is afforded by Newton's concept of absolute time. The following quotation from the Scholium in Book I of the *Principia* is illuminating:

I do not define Time, Space, Place or Motion, as being well known to all. Only I must observe that the vulgar conceive those quantities under no other notions but from the relation they bear to sensible objects. And thence arise certain prejudices, for the removing of which, it will be convenient to distinguish them into Absolute and Relative, True and Apparent, Mathematical and Common.

(1) Absolute, True, and Mathematical Time, of itself, and from its own nature flows equably without regard to anything external, and by another name is called Duration.

Now there is no assurance whatever that there exists in nature anything with properties like those assumed in the definition, and physics, when reduced to concepts of this character, becomes as purely an

abstract science and as far removed from reality as the abstract geometry of the mathematicians, built on postulates. It is a task for experiment to discover whether concepts so defined correspond to anything in nature, and we must always be prepared to find that the concepts correspond to nothing or only partially correspond. In particular, if we examine the definition of absolute time in the light of experiment, we find nothing in nature with such properties.

The new attitude toward a concept is entirely different. We may illustrate by considering the concept of length: what do we mean by the length of an object? We evidently know what we mean by length if we can tell what the length of any and every object is, and for the physicist nothing more is required. To find the length of an object, we have to perform certain physical operations. The concept of length is therefore fixed when the operations by which length is measured are fixed: that is, the concept of length involves as much as and nothing more than the set of operations by which length is determined. In general, we mean by any concept nothing more than a set of operations; *the concept is synonymous with the corresponding set of operations*. If the concept is physical, as of length, the operations are actual physical operations, namely, those by which length is measured; or if the concept is mental, as of mathematical continuity, the operations are mental operations, namely those by which we determine whether a given aggregate of magnitudes is continuous. It is not intended to imply that there is a hard

and fast division between physical and mental concepts, or that one kind of concept does not always contain an element of the other; this classification of concept is not important for our future considerations.

We must demand that the set of operations equivalent to any concept be a unique set, for otherwise there are possibilities of ambiguity in practical applications which we cannot admit.

Applying this idea of "concept" to absolute time, we do not understand the meaning of absolute time unless we can tell how to determine the absolute time of any concrete event, *i.e.*, unless we can measure absolute time. Now we merely have to examine any of the possible operations by which we measure time to see that all such operations are relative operations. Therefore the previous statement that absolute time does not exist is replaced by the statement that absolute time is meaningless. And in making this statement we are not saying something new about nature, but are merely bringing to light implications already contained in the physical operations used in measuring time.

It is evident that if we adopt this point of view toward concepts, namely that the proper definition of a concept is not in terms of its properties but in terms of actual operations, we need run no danger of having to revise our attitude toward nature. For if experience is always described in terms of experience, there must always be correspondence between experience and our description of it, and we need



never be embarrassed, as we were in attempting to find in nature the prototype of Newton's absolute time. Furthermore, if we remember that the operations to which a physical concept are equivalent are actual physical operations, the concepts can be defined only in the range of actual experiment, and are undefined and meaningless in regions as yet untouched by experiment. It follows that strictly speaking we cannot make statements at all about regions as yet untouched, and that when we do make such statements, as we inevitably shall, we are making a conventionalized extrapolation, of the looseness of which we must be fully conscious, and the justification of which is in the experiment of the future.

There probably is no statement either in Einstein or other writers that the change described above in the use of "concept" has been self-consciously made, but that such is the case is proved, I believe, by an examination of the way concepts are now handled by Einstein and others. For of course the true meaning of a term is to be found by observing what a man does with it, not by what he says about it. We may show that this is the actual sense in which concept is coming to be used by examining in particular Einstein's treatment of simultaneity.

Before Einstein, the concept of simultaneity was defined in terms of properties. It was a property of two events, when described with respect to their relation in time, that one event was either before the other, or after it, or simultaneous with it. Simul-

taneity was a property of the two events alone and nothing else; either two events were simultaneous or they were not. The justification for using this term in this way was that it seemed to describe the behavior of actual things. But of course experience then was restricted to a narrow range. When the range of experience was broadened, as by going to high velocities, it was found that the concepts no longer applied, because there was no counterpart in experience for this absolute relation between two events. Einstein now subjected the concept of simultaneity to a critique, which consisted essentially in showing that the operations which enable two events to be described as simultaneous involve measurements on the two events made by an observer, so that "simultaneity" is, therefore, not an absolute property of the two events and nothing else, but must also involve the relation of the events to the observer. Until therefore we have experimental proof to the contrary, we must be prepared to find that the simultaneity of two events depends on their relation to the observer, and in particular on their velocity. Einstein, in thus analyzing what is involved in making a judgment of simultaneity, and in seizing on the act of the observer as the essence of the situation, is actually adopting a new point of view as to what the concepts of physics should be, namely, the operational view.

Of course Einstein actually went much further than this, and found precisely how the operations for judging simultaneity change when the observer moves, and obtained quantitative expressions for the

effect of the motion of the observer on the relative time of two events. We may notice, parenthetically, that there is much freedom of choice in selecting the exact operations; those which Einstein chose were determined by convenience and simplicity with relation to light beams. Entirely apart from the precise quantitative relations of Einstein's theory, however, the important point for us is that if we had adopted the operational point of view, we would, before the discovery of the actual physical facts, have seen that simultaneity is essentially a relative concept, and would have left room in our thinking for the discovery of such effects as were later found.

### *Detailed Discussion of the Concept of Length*

We may now gain further familiarity with the operational attitude toward a concept and some of its implications by examining from this point of view the concept of length. Our task is to find the operations by which we measure the length of any concrete physical object. We begin with objects of our commonest experience, such as a house or a house lot. What we do is sufficiently indicated by the following rough description. We start with a measuring rod, lay it on the object so that one of its ends coincides with one end of the object, mark on the object the position of the other end of the rod, then move the rod along in a straight line extension of its previous position until the first end coincides with the previous position of the second end, repeat this process as often as we can, and call the length the total num-

ber of times the rod was applied. This procedure, apparently so simple, is in practice exceedingly complicated, and doubtless a full description of all the precautions that must be taken would fill a large treatise. We must, for example, be sure that the temperature of the rod is the standard temperature at which its length is defined, or else we must make a correction for it; or we must correct for the gravitational distortion of the rod if we measure a vertical length; or we must be sure that the rod is not a magnet or is not subject to electrical forces. All these precautions would occur to every physicist. But we must also go further and specify all the details by which the rod is moved from one position to the next on the object—its precise path through space and its velocity and acceleration in getting from one position to another. Practically of course, precautions such as these are not mentioned, but the justification is in our experience that variations of procedure of this kind are without effect on the final result. But we always have to recognize that all our experience is subject to error, and that at some time in the future we may have to specify more carefully the acceleration, for example, of the rod in moving from one position to another, if experimental accuracy should be so increased as to show a measureable effect. In *principle* the operations by which length is measured should be *uniquely* specified. If we have more than one set of operations, we have more than one concept, and strictly there should be a separate name to correspond to each different set of operations.

So much for the length of a stationary object, which is complicated enough. Now suppose we have to measure a moving street car. The simplest, and what we may call the "naïve" procedure, is to board the car with our meter stick and repeat the operations we would apply to a stationary body. Notice that this procedure reduces to that already adopted in the limiting case when the velocity of the street car vanishes. But here there may be new questions of detail. How shall we jump on to the car with our stick in hand? Shall we run and jump on from behind, or shall we let it pick us up from in front? Or perhaps does now the material of which the stick is composed make a difference, although previously it did not? All these questions must be answered by experiment. We believe from present evidence that it makes no difference how we jump on to the car, or of what material the rod is made, and that the length of the car found in this way will be the same as if it were at rest. But the experiments are more difficult, and we are not so sure of our conclusions as before. Now there are very obvious limitations to the procedure just given. If the street car is going too fast, we can not board it directly, but must use devices, such as getting on from a moving automobile; and, more important still, there are limitations to the velocity that can be given to street cars or to meter sticks by any practical means in our control, so that the moving bodies which can be measured in this way are restricted to a low range of velocity. If we want to be able to measure the length of bodies mov-



ing with higher velocities such as we find existing in nature (stars or cathode particles), we must adopt another definition and other operations for measuring length, which also reduce to the operations already adopted in the static case. This is precisely what Einstein did. Since Einstein's operations were different from our operations above, *his "length" does not mean the same as our "length."* We must accordingly be prepared to find that the length of a moving body measured by the procedure of Einstein is not the same as that above; this of course is the fact, and the transformation formulas of relativity give the precise connection between the two lengths.

Einstein's procedure for measuring the length of bodies in motion was dictated not only by the consideration that it must be applicable to bodies with high velocities, but also by mathematical convenience, in that Einstein describes the world mathematically by a system of coördinate geometry, and the "length" of an object is connected simply with quantities in the analytic equations.

It is of interest to describe briefly Einstein's actual operations for measuring the length of a body in motion; it will show how operations which may be simple from a mathematical point of view may appear complicated from a physical viewpoint. The observer who is to measure the length of a moving object must first extend over his entire plane of reference (for simplicity the problem is considered two-dimensional) a system of time coördinates, *i.e.*, at each point of his plane of reference there must be a

clock, and all these clocks must be synchronized. At each clock an observer must be situated. Now to find the length of the moving object at a specified instant of time (it is a subject for later investigation to find whether its length is a function of time), the two observers who happen to coincide in position with the two ends of the object at the specified time on their clocks are required to find the distance between their two positions by the procedure for measuring the length of a stationary object, and this distance is by definition the length of the moving object in the given reference system. This procedure for measuring the length of a body in motion hence involves the idea of simultaneity, through the simultaneous position of the two ends of the rod, and we have seen that the operations by which simultaneity are determined are relative, changing when the motion of the system changes. We hence are prepared to find a change in the length of a body when the velocity of the measuring system changes, and this in fact is what happens. The precise numerical dependence is worked out by Einstein, and involves other considerations, in which we are not interested at present.

The two sorts of length, the naïve one and that of Einstein, have certain features in common. In either case in the limit, as the velocity of the measuring system approaches zero, the operations approach those for measuring the length of a stationary object. This, of course, is a requirement in any good definition, imposed by considerations of convenience, and

it is too obvious a matter to need elaboration. Another feature is that the operations equivalent to either concept both involve the motion of the system, so that we must recognize the possibility that the length of a moving object may be a function of its velocity. It is a matter of experiment, unpredictable until tried, that within the limits of present experimental error the naïve length is not affected by motion, and Einstein's length is.

So far, we have extended the concept of length in only one way beyond the range of ordinary experience, namely to high velocities. The extension may obviously be made in other directions. Let us inquire what are the operations by which we measure the length of a very large object. In practice we probably first meet the desirability of a change of procedure in measuring large pieces of land. Here our procedure depends on measurements with a surveyor's theodolite. This involves extending over the surface of the land a system of coördinates, starting from a base line measured with a tape in the conventional way, sighting on distant points from the extremities of the line, and measuring the angles. Now in this extension we have made one very essential change: the angles between the lines connecting distant points are now angles between beams of light. We assume that a beam of light travels in a straight line. Furthermore, we assume in extending our system of triangulation over the surface of the earth that the geometry of light beams is Euclidean. We do the best we can to check the assumptions, but

at most can never get more than a partial check. Thus Gauss<sup>1</sup> checked whether the angles of a large terrestrial triangle add to two right angles and found agreement within experimental error. We now know from the experiments of Michelson<sup>2</sup> that if his measurements had been accurate enough he would not have got a check, but would have had an excess or defect according to the direction in which the beam of light travelled around the triangle with respect to the rotation of the earth. But if the geometry of light beams is Euclidean, then not only must the angles of a triangle add to two right angles, but there are definite relations between the lengths of the sides and the angles, and to check these relations the sides should be measured by the old procedure with a meter stick. Such a check on a large scale has never been attempted, and is not feasible. It seems, then, that our checks on the Euclidean character of optical space are all of restricted character. We have apparently proved that up to a certain scale of magnitude optical space is Euclidean with respect to measures of angle, but this may not necessarily involve that space is also Euclidean with respect to measures of length, so that space need not be completely Euclidean. There is a further most important restriction in that our studies of non-Euclidean geometry have shown that the *percentage* excess of the angles of a non-Euclidean triangle over  $180^\circ$

<sup>1</sup> C. F. Gauss, *Gesammelte Werke*, especially vol. IV.

<sup>2</sup> See a discussion of the theory of this experiment by L. Silberstein, *Jour. Opt. Soc. Amer.* 5, 291-307, 1921.

may depend on the magnitude of the triangle, so that it may well be that we have not detected the non-Euclidean character of space simply because our measurements have not been on a large enough scale.

We thus see that the concept of length has undergone a very essential change of character even within the range of terrestrial measurements, in that we have substituted for what I may call the tactual concept an optical concept, complicated by an assumption about the nature of our geometry. From a very direct concept we have come to a very indirect concept with a most complicated set of operations. Strictly speaking, length when measured in this way by light beams should be called by another name, since the operations are different. The practical justification for retaining the same name is that within our present experimental limits a numerical difference between the results of the two sorts of operations has not been detected.

We are still worse off when we make the extension to solar and stellar distances. Here space is entirely optical in character, and we never have an opportunity of even partially comparing tactual with optical space. No direct measures of length have ever been made, nor can we even measure the three angles of a triangle and so check our assumption that the use of Euclidean geometry in extending the concept of space is justified. We never have under observation more than two angles of a triangle, as when we measure the distance of the moon by observation from the two



ends of the earth's diameter. To extend to still greater distance our measures of length, we have to make still further assumptions, such as that inferences from the Newtonian laws of mechanics are valid. The accuracy of our inferences about lengths from such measurements is not high. Astronomy is usually regarded as a science of extraordinarily high accuracy, but its accuracy is very restricted in character, namely to the measurement of angles. It is probably safe to say that no astronomical distance, except perhaps that of the moon, is known with an accuracy greater than 0.1%. When we push our estimates to distances beyond the confines of the solar system in which we are assisted by the laws of mechanics, we are reduced in the first place to measurements of parallax, which at best have a quite inferior accuracy, and which furthermore fail entirely outside a rather restricted range. For greater stellar distances we are driven to other and much rougher estimates, resting for instance on the extension to great distances of connections found within the range of parallax between brightness and spectral type of a star, or on such assumptions as that, because a group of stars looks as if it were all together in space and had a common origin, it actually is so. Thus at greater and greater distances not only does experimental accuracy become less, but the very nature of the operations by which length is to be determined becomes indefinite, so that the distances of the most remote stellar objects as estimated by different observers or by different methods may be very

divergent. A particular consequence of the inaccuracy of the astronomical measures of great distances is that the question of whether large scale space is Euclidean or not is merely academic.

We thus see that in the extension from terrestrial to great stellar distances the concept of length has changed completely in character. To say that a certain star is  $10^5$  light years distant is actually and conceptually an entire different *kind* of thing from saying that a certain goal post is 100 meters distant. Because of our conviction that the character of our experience may change when the range of phenomena changes, we feel the importance of such a question as whether the space of distances of  $10^5$  light years is Euclidean or not, and are correspondingly dissatisfied that at present there seems no way of giving meaning to it.

We encounter difficulties similar to those above, and are also compelled to modify our procedures, when we go to small distances. Down to the scale of microscopic dimensions a fairly straightforward extension of the ordinary measuring procedure is sufficient, as when we measure a length in a micrometer eyepiece of a microscope. This is of course a combination of tactual and optical measurements, and certain assumptions, justified as far as possible by experience, have to be made about the behavior of light beams. These assumptions are of a quite different character from those which give us concern on the astronomical scale, because here we meet difficulty from interference effects due to the finite

scale of the structure of light, and are not concerned with a possible curvature of light beams in the long reaches of space. Apart from the matter of convenience, we might also measure small distances by the tactual method.

As the dimensions become smaller, certain difficulties become increasingly important that were negligible on a larger scale. In carrying out physically the operations equivalent to our concepts, there are a host of practical precautions to be taken which could be explicitly enumerated with difficulty, but of which nevertheless any practical physicist is conscious. Suppose, for example, we measure length tactually by a combination of Johanssen gauges. In piling these together, we must be sure that they are clean, and are thus in actual contact. Particles of mechanical dirt first engage our attention. Then as we go to smaller dimensions we perhaps have to pay attention to adsorbed films of moisture, then at still smaller dimensions to adsorbed films of gas, until finally we have to work in a vacuum, which must be the more nearly complete the smaller the dimensions. About the time that we discover the necessity for a complete vacuum, we discover that the gauges themselves are atomic in structure, that they have no definite boundaries, and therefore no definite length, but that the length is a hazy thing, varying rapidly in time between certain limits. We treat this situation as best we can by taking a time average of the apparent positions of the boundaries, assuming that along with the decrease of dimensions we have ac-

quired a corresponding extravagant increase in nimbleness. But as the dimensions get smaller continually, the difficulties due to this haziness increase indefinitely in percentage effect, and we are eventually driven to give up altogether. We have made the discovery that there are *essential* physical limitations to the operations which defined the concept of length. [We perhaps do not regard the substitution of optical for tactual space on the astronomical scale as compelled by the same sort of physical necessity, because I suppose the possible eventual landing of men in the moon will always be one of the dreams of humanity.] At the same time that we have come to the end of our rope with our Johanssen gauge procedure, our companion with the microscope has been encountering difficulties due to the finite wave length of light; this difficulty he has been able to minimize by using light of progressively shorter wave lengths, but he has eventually had to stop on reaching X-rays. Of course this optical procedure with the microscope is more convenient, and is therefore adopted in practice.

Let us now see what is implied in our concept of length extended to ultramicroscopic dimensions. What, for instance, is the meaning of the statement that the distance between the planes of atoms in a certain crystal is  $3 \times 10^{-8}$  cm.? What we would like to mean is that  $1/3 \times 10^8$  of these planes piled on top of each other give a thickness of 1 cm.; but of course such a meaning is not the actual one. The actual meaning may be found by examining the operations

by which we arrived at the number  $3 \times 10^{-8}$ . As a matter of fact,  $3 \times 10^{-8}$  was the number obtained by solving a general equation derived from the wave theory of light, into which certain numerical data obtained by experiments with X-rays had been substituted. Thus not only has the character of the concept of length changed from tactual to optical, but we have gone much further in committing ourselves to a definite optical theory. If this were the whole story, we would be most uncomfortable with respect to this branch of physics, because we are so uncertain of the correctness of our optical theories, but actually a number of checks can be applied which greatly restore our confidence. For instance, from the density of the crystal and the grating space, the weight of the individual atoms may be computed, and these weights may then be combined with measurements of the dimensions of other sorts of crystal into which the same atoms enter to give values of the densities of these crystals, which may be checked against experiment. All such checks have succeeded within limits of accuracy which are fairly high. It is important to notice that, in spite of the checks, the character of the concept is changing, and begins to involve such things as the equations of optics and the assumption of the conservation of mass.

We are not content, however, to stop with dimensions of atomic order, but have to push on to the electron with a diameter of the order of  $10^{-13}$  cm. What is the possible meaning of the statement that the diameter of an electron is  $10^{-13}$  cm.? Again the only an-



swer is found by examining the operations by which the number  $10^{-12}$  was obtained. This number came by solving certain equations derived from the field equations of electrodynamics, into which certain numerical data obtained by experiment had been substituted. The concept of length has therefore now been so modified as to include that theory of electricity embodied in the field equations, and, most important, assumes the correctness of extending these equations from the dimensions in which they may be verified experimentally into a region in which their correctness is one of the most important and problematical of present-day questions in physics. To find whether the field equations are correct on a small scale, we must verify the relations demanded by the equations between the electric and magnetic forces and the space coördinates, to determine which involves measurement of lengths. But if these space coördinates cannot be given an independent meaning apart from the equations, not only is the attempted verification of the equations impossible, but the question itself is meaningless. If we stick to the concept of length by itself, we are landed in a vicious circle. As a matter of fact, the concept of length disappears as an independent thing, and fuses in a complicated way with other concepts, all of which are themselves altered thereby, with the result that the total number of concepts used in describing nature at this level is reduced in number. A precise analysis of the situation is difficult, and I suppose has never been attempted, but the general character of the situation is

evident. Until at least a partial analysis is attempted, I do not see how any meaning can be attached to such questions as whether space is Euclidean in the small scale.

It is interesting to observe that any increased accuracy in knowledge of large scale phenomena must, as far as we now can see, arise from an increase in the accuracy of measurement of small things, that is, in the measurement of small angles or the analysis of minute differences of wave lengths in the spectra. To know the very large takes us into the same field of experiment as to know the very small, so that operationally the large and the small have features in common.

This somewhat detailed analysis of the concept of length brings out features common to all our concepts. If we deal with phenomena outside the domain in which we originally defined our concepts, we may find physical hindrances to performing the operations of the original definition, so that the original operations have to be replaced by others. These new operations are, of course, to be so chosen that they give, within experimental error, the same numerical results in the domain in which the two sets of operations may be both applied; but we must recognize in principle that in changing the operations we have really changed the concept, and that to use the same name for these different concepts over the entire range is dictated only by considerations of convenience, which may sometimes prove to have been purchased at too high a price in terms of unambiguity.

We must always be prepared some day to find that an increase in experimental accuracy may show that the two different sets of operations which give the same results in the more ordinary part of the domain of experience, lead to measurably different results in the more unfamiliar parts of the domain. We must remain aware of these joints in our conceptual structure if we hope to render unnecessary the services of the unborn Einsteins.

The second feature common to all concepts brought out by the detailed discussion of length is that, as we approach the experimentally attainable limit, concepts lose their individuality, fuse together, and become fewer in number, as we have seen that at dimensions of the order of the diameter of an electron the concepts of length and the electric field vectors fuse into an amorphous whole. Not only does nature as experienced by us become different in character on its horizons, but it becomes simpler, and therefore our concepts, which are the building stones of our descriptions, become fewer in number. This seems to be an entirely natural state of affairs. How the number of concepts is often kept formally the same as we approach the horizon will be discussed later in special cases.

A precise analysis of our conceptual structure has never been attempted, except perhaps in very restricted domains, and it seems to me that there is room here for much important future work. Such an analysis is not to be attempted in this essay, but only some of the more important qualitative aspects are to

be pointed out. It will never be possible to give a clean-cut logical analysis of the conceptual situation, for the nature of our concepts, according to our operational point of view, is the same as the nature of experimental knowledge, which is often hazy. Thus in the transition regions where nature is getting simpler and the number of operationally independent concepts changes, a certain haziness is inevitable, for the actual change in our conceptual structure in these transition regions is continuous, corresponding to the continuity of our experimental knowledge, whereas formally the number of concepts should be an integer.

### *The Relative Character of Knowledge*

Two other consequences of the operational point of view must now be examined. First is the consequence that all our knowledge is relative. This may be understood in a general or a more particular sense. The general sense is illustrated in Haldane's book on the *Reign of Relativity*. Relativity in the general sense is the merest truism if the operational definition of concept is accepted, for experience is described in terms of concepts, and since our concepts are constructed of operations, all our knowledge must unescapably be relative to the operations selected. But knowledge is also relative in a narrower sense, as when we say there is no such thing as absolute rest (or motion) or absolute size, but rest and size are relative terms. Conclusions of this kind are involved in the specific character of the operations in

terms of which rest or size are defined. An examination of the operations by which we determine whether a body is at rest or in motion shows that the operations are relative operations: rest or motion is determined with respect to some other body selected as the standard. In saying that there is no such thing as absolute rest or motion we are not making a statement about nature in the sense that might be supposed, but we are merely making a statement about the character of our descriptive processes. Similarly with regard to size: examination of the operations of the measuring process shows that size is measured relative to the fundamental measuring rod.

The "absolute" therefore disappears in the original meaning of the word. But the "absolute" may usefully return with an altered meaning, and we may say that a thing has absolute properties if the numerical magnitude is the same when measured with the same formal procedure by all observers. Whether a given property is absolute or not can be determined only by experiment, landing us in the paradoxical position that the absolute is absolute only relative to experiment. In some cases, the most superficial observation shows that a property is not absolute, as, for example, it is at once obvious that measured velocity changes with the motion of the observer. But in other cases the decision is more difficult. Thus Michelson thought he had an absolute procedure for measuring length, by referring to the wave length of



the red cadmium line as standard;<sup>1</sup> it required difficult and accurate experiment to show that this length varies with the motion of the observer. Even then, by changing the definition of the length of a moving object, we believe that length might be made to re-assume its desired absolute character.

To stop the discussion at this point might leave the impression that this observation of the relative character of knowledge is of only a very tenuous and academic interest, since it appears to be concerned mostly with the character of our descriptive processes, and to say little about external nature. [What this means we leave to the metaphysician to decide.] But I believe there is a deeper significance to all this. It must be remembered that all our argument starts with the concepts as given. Now these concepts involve physical operations; in the discovery of what operations may be usefully employed in describing nature is buried almost all physical experience. In erecting our structure of physical science, we are building on the work of all the ages. There is then this purely physical significance in the statement that all motion is relative, namely that no operations of measuring motion have been found to be useful in describing simply the behavior of nature which are not operations relative to a single observer; in making this statement we are stating something about nature. It takes an enormous amount of real physical experi-

<sup>1</sup> A. A. Michelson, *Light Waves and Their Uses*, University of Chicago Press, 1903, Chap. V.

ence to discover relations of this sort. The discovery that the number obtained by counting the number of times a stick may be applied to an object can be simply used in describing natural phenomena was one of the most important and fundamental discoveries ever made by man.

### *Meaningless Questions*

Another consequence of the operational character of our concepts, almost a corollary of that considered above, is that it is quite possible, nay even disquietingly easy, to invent expressions or to ask questions that are meaningless. It constitutes a great advance in our critical attitude toward nature to realize that a great many of the questions that we uncritically ask are without meaning. If a specific question has meaning, it must be possible to find operations by which an answer may be given to it. It will be found in many cases that the operations cannot exist, and the question therefore has no meaning. For instance, it means nothing to ask whether a star is at rest or not. Another example is a question proposed by Clifford, namely, whether it is not possible that as the solar system moves from one part of space to another the absolute scale of magnitude may be changing, but in such a way as to affect all things equally, so that the change of scale can never be detected. An examination of the operations by which length is measured in terms of measuring rods shows that the operations do not exist (because of the nature of our definition of length) for answering the ques-

tion. The question can be given meaning only from the point of view of some imaginary superior being watching from an external point of vantage. But the operations by which such a being measures length are different from the operations of our definition of length, so that the question acquires meaning only by changing the significance of our terms—in the original sense the question means nothing.

To state that a certain question about nature is meaningless is to make a significant statement about nature itself, because the fundamental operations are determined by nature, and to state that nature cannot be described in terms of certain operations is a significant statement.

It must be recognized, however, that there is a sense in which no serious question is entirely without meaning, because doubtless the questioner had in mind some intention in asking the question. But to give meaning in this sense to a question, one must inquire into the meaning of the concepts as used by the questioner, and it will often be found that these concepts can be defined only in terms of fictitious properties, as Newton's absolute time was defined by its properties, so that the meaning to be ascribed to the question in this way has no connection with reality. I believe that it will enable us to make more significant and interesting statements, and therefore will be more useful, to adopt exclusively the operational view, and so admit the possibility of questions entirely without meaning.

This matter of meaningless questions is a very

subtle thing which may poison much more of our thought than that dealing with purely physical phenomena. I believe that many of the questions asked about social and philosophical subjects will be found to be meaningless when examined from the point of view of operations. It would doubtless conduce greatly to clarity of thought if the operational mode of thinking were adopted in all fields of inquiry as well as in the physical. Just as in the physical domain, so in other domains, one is making a significant statement about his subject in stating that a certain question is meaningless.

In order to emphasize this matter of meaningless questions, I give here a list of questions, with which the reader may amuse himself by finding whether they have meaning or not.

- (1) Was there ever a time when matter did not exist?
- (2) May time have a beginning or an end?
- (3) Why does time flow?
- (4) May space be bounded?
- (5) May space or time be discontinuous?
- (6) May space have a fourth dimension, not directly detectible, but given indirectly by inference?
- (7) Are there parts of nature forever beyond our detection?
- (8) Is the sensation which I call blue really the *same* as that which my neighbor calls blue? Is it possible that a blue object may arouse in him the same sensation that a red object does in me and *vice versa*?
- (9) May there be missing integers in the series of natural numbers as we know them?

- (10) Is a universe possible in which  $2+2 \neq 4$ ?
- (11) Why does negative electricity attract positive?
- (12) Why does nature obey laws?
- (13) Is a universe possible in which the laws are different?
- (14) If one part of our universe could be *completely* isolated from the rest, would it continue to obey the same laws?
- (15) Can we be sure that our logical processes are valid?

### GENERAL COMMENTS ON THE OPERATIONAL POINT OF VIEW

To adopt the operational point of view involves much more than a mere restriction of the sense in which we understand "concept," but means a far-reaching change in all our habits of thought, in that we shall no longer permit ourselves to use as tools in our thinking concepts of which we cannot give an adequate account in terms of operations. In some respects thinking becomes simpler, because certain old generalizations and idealizations become incapable of use; for instance, many of the speculations of the early natural philosophers become simply unreadable. In other respects, however, thinking becomes much more difficult, because the operational implications of a concept are often very involved. For example, it is most difficult to grasp adequately all that is contained in the apparently simple concept of "time," and requires the continual correction of mental tendencies which we have long unquestioningly accepted.



Operational thinking will at first prove to be an unsocial virtue; one will find oneself perpetually unable to understand the simplest conversation of one's friends, and will make oneself universally unpopular by demanding the meaning of apparently the simplest terms of every argument. Possibly after every one has schooled himself to this better way, there will remain a permanent unsocial tendency, because doubtless much of our present conversation will then become unnecessary. The socially optimistic may venture to hope, however, that the ultimate effect will be to release one's energies for more stimulating and interesting interchange of ideas.

Not only will operational thinking reform the social art of conversation, but all our social relations will be liable to reform. Let any one examine in operational terms any popular present-day discussion of religious or moral questions to realize the magnitude of the reformation awaiting us. Whenever we temporize or compromise in applying our theories of conduct to practical life we may suspect a failure of operational thinking.

## CHAPTER II

### OTHER GENERAL CONSIDERATIONS

#### THE APPROXIMATE CHARACTER OF EMPIRICAL KNOWLEDGE

ALTHOUGH many aspects of the processes by which we obtain knowledge of the external physical world are much beyond the scope of our present inquiry, one matter must be mentioned in detail because it tacitly underlies all our discussion, the fact, namely, that all results of measurement are only approximate. That such is true is evident after the most superficial examination of any measuring process; any statement about numerical relations between measured quantities must always be subject to the qualification that the relation is valid only within limits. Furthermore, all experience seems to be of this character; we never have perfectly clean-cut knowledge of anything, but all our experience is surrounded by a twilight zone, a penumbra of uncertainty, into which we have not yet penetrated. This penumbra is as truly an unexplored region as any other region beyond experiment, such as the region of high velocities, for example, and we must hold no preconceived notions as to what will be found within

the region. The penumbra is to be penetrated by improving the accuracy of measurement. Within what was at one time penumbra has been found the displacement of angular position of the stars near the edge of the solar disc, and within the penumbra as yet unpenetrated we look for such effects as the equivalence of mass and energy. Many of the great discoveries of the future will probably be made within the penumbra: we have already mentioned that increased knowledge of phenomena of a cosmic scale is to be obtained by increasing the accuracy of measurement of the very small.

It is a general consequence of the approximate character of all measurement that no empirical science can ever make exact statements. This was fairly obvious in the case of mechanics, but it required a Gauss<sup>1</sup> to convince us that the geometry in which we are interested as physicists is an empirical subject, and that one cannot say that actual space is Euclidean, but only that actual space approaches to ideal Euclidean space within a certain degree of approximation. I believe that we are compelled to go still further, and recognize that arithmetic, so far as it purports to deal with actual physical objects, is also affected with the same penumbra of uncertainty as all other empirical science. A typical statement of empirical arithmetic is that 2 objects plus 2 objects makes 4 objects. This statement acquires physical meaning only in terms of certain physical operations, and these operations must be performed in

<sup>1</sup> C. F. Gauss, *Gesammelte Werke*, especially vols. IV and VIII.

time. Now the penumbra gets into this situation through the concept of object. If the statement of arithmetic is to be an exact statement in the mathematical sense the "object" must be a definite clean-cut thing, which preserves its identity in time with no penumbra. But this sort of thing is never experienced, and as far as we know does not correspond exactly to anything in experience. It is of course true that in most experience the penumbra is so very thin and snug-fitting that it requires special effort to recognize its presence at all; but scrutiny, I believe, shows that it is always there. If our experience had been restricted to phenomena in a vacuum, and the objects we were trying to count had been spheres of a gas which expand and interpenetrate, it is obvious that the concept of "object" as a thing with identity would have been much more difficult to form. Or, if our objects are tumblers of water, we discover when our observation reaches a certain stage of refinement that the amount of water is continually changing by evaporation and condensation, and we are bothered by the question whether the object is still the same after it has waxed and waned. Coming to solids, we eventually discover that even solids evaporate, or condense gases on them, and we see that an object with identity is an abstraction corresponding exactly to nothing in nature. Of course the penumbra of uncertainty which surrounds our arithmetical statements because of this property of physical objects is so exceedingly tenuous that practically we are not aware of its existence, and expect never to

find undiscovered phenomena within the penumbra. But in principle we must recognize its presence, and must further recognize that *all* empirical science must be of this character.

In most empirical sciences, the penumbra is at first prominent, and becomes less important and thinner as the accuracy of physical measurement is increased. In mechanics, for example, the penumbra is at first like a thick obscuring veil at the stage where we measure forces only by our muscular sensations, and gradually is attenuated, as the precision of measurements increases. But with the arithmetical concept of an individual identifiable object it is the exact reverse; a crude point of view does not suspect the existence of the penumbra at all, and we discover it only by highly refining our methods. Doubtless arithmetic owes its early development to this property.

We may now go still further. Operations themselves are, of course, derived from experience, and would be expected also to have a nebulous edge of uncertainty. We have to ask such questions as whether the *operations* of arithmetic are clean-cut things. Is the operation of multiplying 2 objects by 2 a definite operation, with no enveloping haze? All our physical experience convinces us that if there is a penumbra about the concept of operations of this sort it is so tenuous as to be negligible, at least for the present; but the question affords an interesting topic for speculation. We also have to ask whether mental operations may similarly be enveloped in a haze.



## EXPLANATIONS AND MECHANISMS

Perhaps the climax of our task of interpreting and correlating nature is reached when we are able to find an explanation of phenomena; with the finding of the explanation we are inclined to feel that our understanding of the situation is complete. We now have to ask what is the nature of the explanation which we set as the goal of our efforts. The answer is not easy to give, and there may be difference of opinion about it. We shall get the best answer to this, as to so many other questions, by adopting the operational point of view, and examining what we do in giving an explanation. I believe that examination will show that the essence of an explanation consists in reducing a situation to elements with which we are so familiar that we accept them as a matter of course, so that our curiosity rests.<sup>1</sup> "Reducing a situation to elements" means, from the operational point of view, discovering familiar correlations between the phenomena of which the situation is composed.

There is involved here the thesis that it is possible to analyze nature into correlations, without, however, any assumption whatever as to the character of these correlations. It seems to me that such a thesis is the most general that can be made if nature is to be intelligible at all. This thesis underlies all the considerations of this essay, and we shall not try to find anything more general. We shall, however,

<sup>1</sup> The ultimate elements of explanation are analogous to the axioms of formal mathematics.

recognize that any assumption as to the character of the correlations constitutes a special hypothesis which may restrict the future, and that therefore these special hypotheses are to be subjected to special examination. We return to this matter in more detail in discussing the causality concept, which is closely related to the concept of explanation.

In this view of explanation there is no implication that the "element" is either a smaller or a larger scale thing than the phenomenon being explained; thus we may explain the properties of a gas in terms of its constituent molecules, or perhaps some day we shall become so familiar with the idea of a non-Euclidean space that we shall *explain* (instead of describe) the gravitational attraction of a stone by the earth in terms of a space-time curvature imposed by all the rest of the matter in the universe.

If this is accepted as the true nature of explanation, we see that an explanation is not an absolute sort of thing, but what is satisfactory for one man will not be for another. The savage is satisfied by explaining the thunderstorm as the capricious act of an angry god. The physicist demands more, and requires that the familiar elements to which we reduce a situation be such that we can intuitively predict their behavior. Thus even if the physicist believed in the existence of the angry god, he would not be satisfied with this explanation of the thunderstorm because he is not so well acquainted with angry gods as to be able to predict when anger is followed by a storm. He would have to know why the god had become angry, and

why making a thunderstorm eased his ire. But even with this additional qualification, scientific explanation is obviously still a relative affair—relative to the elements or axioms to which we make reduction and which we accept as ultimate. These elements depend to a certain extent on the purpose in view, and also on the range of our previous physical experience. If we are explaining the action of a machine, we are satisfied to reduce the action to the push and pull of the various members of the machine, it being accepted as an ultimate that these members transmit pushes or pulls. But the physicist who has extended his experimental knowledge further, may want to explain how the members transmit pushes or pulls in terms of the action on each other of the electrons in their orbits in the atoms. The character of our explanatory structure will depend on the character of our experimental knowledge, and will change as this changes.

Formally, there is no limit to the process of explanation, because we can always ask what is the explanation of the elements in terms of which we have given the last explanation. But the point of view of operations shows that this is mere formalism which ends only in meaningless jargon, for we soon arrive at the limit of our experimental knowledge, and beyond this the operations involved in the concepts of our explanations become impossible and the concepts become meaningless.

As we extend experimental knowledge and push our explanations further and further, we see that the

explanatory sequence may be terminated in several possible ways. In the first place, we may never push our experiments beyond a stage into which the elements with which we are already familiar do not enter. In this case explanation is very simple: it involves nothing essentially new, but merely the disentanglement of complexities. The kinetic theory of gases, in explaining the thermal properties of a gas in terms of ordinary mechanical properties of the molecules, suggests such a situation. Or, secondly, our experiments may bring us into contact with situations novel to us, in which we can recognize no familiar elements, or at least must recognize that there is something in addition to the familiar elements. Such a situation constitutes an explanatory crisis and explanation has to stop by definition. Or thirdly, we may try to force our explanations into a predetermined mold, by formally erecting or inventing beyond the range of present experiment ultimates more or less like elements already familiar to us, and seek to explain all present experience in terms of these chosen ultimates.

Leaving for the present the third possibility, which is within our control to accept or reject, and is a formal matter, it is merely a question of experimental fact which of the first two possibilities corresponds to the actual state of affairs. The most perfunctory examination of the present state of physics shows that we are now facing the second of these possibilities, and that in the new experimental facts of relativity, and in all quantum phenomena, we are confronted

with an explanatory crisis. It has often been emphasized that Einstein's theory of gravitation does not seek at all to give an explanation of gravitational phenomena, but merely describes and correlates these phenomena in comparatively simple mathematical language. No more attempt is made to reduce the gravitational attraction between the earth and the sun to simple terms than was made by Newton. In the realm of quantum phenomena it is of course the merest commonplace that our old ideas of mechanics and electrodynamics have failed, so that it is a matter of the greatest concern to find how many, or indeed whether any, elements of the old situations can be carried over into the new.

An examination of many of the so-called "explanations" of quantum theory constitutes at once a justification of the definition of explanation given above, and of the statement that in quantum phenomena we are at an explanatory crisis. For the endeavor of all these quantum explanations is to find in every new or more complicated situation the same elements which have already been met in simpler situations, and which are therefore relatively more familiar. For example, many quantum phenomena are made to involve the emission of energy when an electron jumps from one orbit to another. But always the elements to which reduction is made are themselves quantum phenomena, and these are still so new and unfamiliar that we feel an instinctive need for explanation in other terms. We seek to understand why the electron emits when it jumps.



The explanatory crisis which now confronts us in relativity and quantum phenomena is but a repetition of what has occurred many time in the past. A similar crisis confronted Prometheus when he discovered fire, and the first man who observed a straw sticking to a piece of rubbed amber, or a suspended lodestone seeking the north star. Every kitten is confronted with such a crisis at the end of nine days. Whenever experience takes us into new and unfamiliar realms, we are to be at least prepared for a new crisis.

Now what are we to do in such a crisis? It seems to me that the only sensible course is to do exactly what the kitten does, namely, to wait until we have amassed so much experience of the new kind that it is perfectly familiar to us, and then to resume the process of explanation with elements from our new experience included in our list of axioms. Not only will observation show that this is what is now actually being done with respect to quantum and gravitational phenomena, but it is in harmony with the entire spirit of our outlook on nature. All our knowledge is in terms of experience; we should not expect or desire to erect an explanatory structure different in character from that of experience. Our experience is finite; on the confines of the experimentally attainable it becomes hazy, and the concepts in terms of which we describe it fuse together and lose independent meaning. Furthermore, at every extension of our experimental range we must be prepared to find, and as a matter of fact we have often found,

that we encounter phenomena of an entirely novel character for which previous experience has given us no preparation. The explanatory structure proposed above has all these properties; it is finite, being terminated by the edge of experiment, the final stages of our explanations are hazy in that it becomes more and more difficult to distinguish elements of familiar experience, and every now and then we must admit new elements into our explanations.

The first step in resuming our explanatory progress, after we have been confronted with such a crisis, is to seek for various sorts of correlation between the elements of our new experience, in the confident expectation that these elements will eventually become so familiar to us that they may be used as the ultimates of a new explanation. This is exactly what is now happening in quantum theory.

Diametrically opposed to the views above, there is another ideal of the explanatory process which is held by many physicists, and which has been mentioned above as the third possible way in which the explanatory sequence may be terminated, namely, the endeavor to devise beyond the limit of present experiment a structure built of elements like some of those of our present experience, in the action of which we endeavor to find the explanation of phenomena in the present range. Now a program such as this, as a serious program for the final correlation of nature, is entirely opposed to the spirit of the considerations expounded here. There is no warrant whatever in experience for the conviction that as we

penetrate deeper and deeper we shall find the elements of previous experience repeated, although sometimes we do find such repetitions, as in the behavior of gases. Yet this has been the attitude of many eminent physicists, for example, Faraday and Maxwell, in seeking to explain distant electrical action by the propagation through a medium of a mechanical push or pull, or by Hertz, who sought in all phenomena the effect of concealed masses with ordinary mechanical inertia. Although as a general principle this program seems to be absolutely without justification, nevertheless it may be justified if the specific character of the physical facts seems to indicate a repetition at lower levels of elements familiar higher up. Hertz undoubtedly had this justification, as did also Maxwell to a certain extent, in the discovery that the general equations of electrodynamics are of the same form as the generalized Lagrangean equations of mechanics. For Faraday, however, there seems no such justification; the urge to this sort of thing in Faraday came from an uncritical acceptance of his own temperamental reactions.

From a less serious point of view it may, however, be quite justified to make such a working hypothesis as that in the action of electrical forces may be discovered the same elements with which we are familiar in the everyday experiences of mechanics. For such a hypothesis often enables us to make partial correlations which suggest new experimental tests, and thus gives the stimulus to an extension of our experimental horizon. Many physicists recog-

nize the tentative character of such attempted explanations, but others apparently take them more seriously, as for example Lord Kelvin in his continuous life-long attempts to find a mechanical explanation of all physical phenomena. This quotation from Kelvin is illuminating. "I never satisfy myself until I can make a mechanical model of a thing. If I can make a mechanical model, I can understand it. As long as I cannot make a mechanical model all the way through, I cannot understand it. . . . But I want to understand light as well as I can without introducing things that we understand even less of."

So much for general considerations on the nature of explanation. Coming now to greater detail, many explanations involve what may be described as a mechanism. It is difficult to characterize exactly what we mean by mechanism, but it seems to be associated with an attitude of mind that strives to realize the third of the possibilities mentioned above. As a matter of fact, the mechanism sought for is usually of a particular type, in that the ultimate elements selected are mechanical elements. This point of view is particularly characteristic of the English school of physicists. Although "mechanism" usually implies mechanical elements, we may show by specific examples that we do actually use the word in a broader significance. If, for example, we could devise within the core of an atom a revolving system of electrical charges, acting on each other with the ordinary inverse square forces of electrostatics, such that

every now and then the system becomes unstable and breaks up, we should doubtless say that we had found a mechanism for explaining radioactive disintegration.

However, the formulation of a precise definition of mechanism is of secondary concern to us; we are primarily interested in understanding the attitude of mind that feels a mechanism is necessary. A typical example of such an urge to a mechanism is afforded by the gravitational action between distant bodies. To many minds the concept of action at a distance is absolutely abhorrent, not to be tolerated for an instant. Such an intolerable situation is avoided by the invention of a medium filling all space, which transmits a force from one body to the other through the successive action on each other of its contiguous parts. Or the dilemma of action at a distance may be avoided in other ways, as by Boscovitch in the eighteenth century, who, in order to explain gravitation, filled space with a triply infinite horde of infinitesimal projectiles. Now of course it is a matter for experiment to decide whether any physical reality can be ascribed to a medium which makes gravitation possible by the action of its adjacent parts, but I can see no justification whatever for the attitude which refuses on purely *a priori* grounds to accept action at a distance as a possible axiom or ultimate of explanation. It is difficult to conceive anything more scientifically bigoted than to postulate that all possible experience conforms to the same type as that with which we are already familiar, and



therefore to demand that explanation use only elements familiar in everyday experience. Such an attitude bespeaks an unimaginativeness, a mental obtuseness and obstinacy, which might be expected to have exhausted their pragmatic justification at a lower plane of mental activity.

Although it will probably be fairly easy to give intellectual assent to the strictures of the last paragraph, I believe many will discover in themselves a longing for mechanical explanation which has all the tenacity of original sin. The discovery of such a desire need not occasion any particular alarm, because it is easy to see how the demand for this sort of explanation has had its origin in the enormous preponderance of the mechanical in our physical experience. But nevertheless, just as the old monks struggled to subdue the flesh, so must the physicist struggle to subdue this sometimes nearly irresistible, but perfectly unjustifiable desire. One of the large purposes of this exposition will be attained if it carries the conviction that this longing is unjustifiable, and is worth making the effort to subdue.

The situation with respect to action at a distance is typical of the general situation. I believe the essence of the explanatory process is such that we must be prepared to accept as an ultimate for our explanations the mere statement of a correlation between phenomena or situations with which we are sufficiently familiar. Thus, in quantum theory, there is no reason why we should not be willing to accept as an ultimate the fundamental fact that when an elec-

tron jumps radiation is emitted, provided always that we can give independent meaning in terms of operations to the jumping of an electron. If there is no experiment suggesting other and intermediate phenomena, we ought to be able to rest intellectually satisfied with this. Of course it is quite a different matter, and entirely justified, to imagine what the assumption of finer details in the process would involve experimentally, and then to seek for these possible new experimental facts.

It is a consequence of this view that any correlation is adapted to be an absolutely final element of explanation, and can never be superseded by the discovery of new experimental facts, if the correlation is *by definition* beyond the reach of further experiment. Such a possibility, for example, is contained in a correlation between the numerical magnitude of the gravitational constant and the total mass of the universe. Something of this sort may be well attempted by those who desire their explanations to take a formally final shape. We shall return to this subject later.

The instinctive demand for a mechanism is fortified by observation of the many important cases in which mechanisms have been discovered or invented. However, the significance of such successful attempts must be subject to most careful scrutiny. The matter has been discussed by Poincaré,<sup>1</sup> who showed that

<sup>1</sup> Henri Poincaré. *Wissenschaft and Hypothese*, Translated into German by F. and L. Lindemann, Teubner, Leipzig, 1906. See especially p. 217.

not only is it always possible to find a mechanistic explanation of any phenomenon (Hertz's program was a perfectly possible one), but there are always an infinite number of such explanations. This is very unsatisfactory. We want to be able to find the *real* mechanism. Now an examination of specific proposed mechanisms will show that most mechanisms are more complicated than the simple physical phenomenon which they are invented to explain, in that they have more independently variable attributes than the phenomenon has been yet proved to have. An example is afforded by the mechanical models invented to facilitate the study of the properties of simple inductive electrical circuits. The great number of such models which have been proposed is sufficient indication of their possible infinite number. But if the mechanism has more independently variable attributes than the original phenomenon, it is obvious that the question is without meaning whether the mechanism is the real one or not, for in the mechanism there must be simple motions or combinations of motions which have no counterpart in features of the original phenomenon as yet discovered. Obviously, then, the operations do not exist by which we may set up a one to one correspondence between the properties of the mechanism and the natural phenomenon, and the question of reality has no meaning. If, then, a mechanism is to be taken seriously as actually corresponding to reality, we must demand that it have no more degrees of freedom than the original phenomenon, and we must also be sure that

the phenomenon has no undiscovered features. Physical experience shows that such conditions are most difficult to meet, and indeed the probability is that they are impossible.

A mechanism with more independently variable attributes than the phenomenon may prove to be a very useful tool of thought, and therefore worth inventing and studying, but it is to be regarded no more seriously than is a mnemonic device, or any other artifice by which a man forces his mind to give him better service.

There is another possible program of explanation, the converse of that considered above, namely, to explain all familiar facts of ordinary experience in terms of less familiar facts found at a deeper level. The most striking example of this is the recent attempt to give a complete electrical explanation of the universe. The original attempt was to explain electrical effects in mechanical terms; this attempt failed. At about the same time the existence of the electron was experimentally established, so that it was evident that electricity is a very fundamental constituent of matter. The program of explanation was reversed, and an electrical explanation sought for all mechanical phenomena, including in particular mechanical mass. But this attempt has also failed; we recognize that part of mass may be non-electrical in character, we postulate non-electrical forces inside the electron, and further, we usually postulate for electrons and protons the property of impenetrability, a property derived from experience on a higher scale of magnitude.

A program of this general sort is likely to be regarded with considerable sympathy, and indeed the chances of success seem much greater than do those of the converse program, for in our experience large scale phenomena are more often built up from and analyzed into small scale phenomena than the converse. But as a matter of principle we must again recognize that the only appeal is to experiment, and that we have to ask just one question: "Is it true, as a matter of fact, that all large scale phenomena can be built up of elements of small scale phenomena?" It seems to me that the experimental warrant for this conviction has not yet been given. The failure of the attempted electrical explanation of the universe is a case in point. However, the failure to prove a proposition is no guarantee that some time it may not be proved, and many physicists are convinced of the ultimate feasibility of this program. Personally I feel that the large may not always be analyzed into the smaller; the subject will be discussed again.

A conviction of the significance of microscopic analysis has many features in common with the usual conviction of the ultimate simplicity of nature. The thesis of simplicity involves in addition the assumption that the kinds of small scale elements are few in number, but actually this involves no important difference between the two convictions, because we have seen that the elements of which we build our structure become fewer in number as we approach the limit of the experimentally attainable. We may properly grant to convictions of this sort pragmatic value in suggesting new correlations and experiments,



but a recognition of the empirical basis of all physics will not allow us to go further.

### MODELS AND CONSTRUCTS

In discussing the concept of length, we could find no meaning in questions such as: "Is space on a scale of  $10^{-8}$  cm. Euclidean?" Nevertheless it will seem to many that they do attach a perfectly definite meaning to a question of this kind. Of course it must be agreed that magnitudes of  $10^{-8}$  cm. cannot be thought of in terms of immediate sensation. When one thinks of an atom as a thing with any geometrical properties at all, I believe he will find that what he essentially does is to imagine a model, multiplying all the hypothetical dimensions by a factor large enough to bring it to a magnitude of ordinary experience. This large scale model is given properties corresponding to those of the physical thing. For example, the model of the atom which was accepted in the fall of 1925 contains electrons rotating in orbits, and every now and then an electron jumps from one orbit to another, and simultaneously energy is radiated from the atom. Such a model is satisfactory if it offers the counterpart of all the phenomena of the original atom. Now I believe the only meaning that any one can find in his statement that the space of the atom is Euclidean is that he believes that he can construct in Euclidean space a model with all the observed properties of the atom. This possibility may or may not be sufficient to give real physical significance to the statement that the space of the atom

is Euclidean. The situation here is very much the same as it was with respect to mechanisms. The model may have many more properties than correspond to measurable properties of the atom, and in particular, the operations by which the space of the model is tested for its Euclidean character may [and as a matter of fact I believe *do*] not have any counterpart in operations which can be carried out on the atom. Further, we cannot attach any *real* significance to the statement that the space of the atom is Euclidean unless we can show that no model constructed in non-Euclidean space can reproduce the measurable properties of the atom.

In spite of all this, I believe that the model is a useful and indeed unescapable tool of thought, in that it enables us to think about the unfamiliar in terms of the familiar. There are, however, dangers in its use: it is the function of criticism to disclose these dangers, so that the tool may be used with confidence.

Closely related to the mental model are mental constructs, of which physics is full. There are many sorts of constructs: those in which we are interested are made by us to enable us to deal with physical situations which we cannot directly experience through our senses, but with which we have contact indirectly and by inference. Such constructs usually involve the element of invention to a greater or less degree. A construct containing very little of invention is that of the inside of an opaque solid body. We can never experience directly through our senses the inside of such a solid body, because the instant we directly

experience it, it ceases by definition to be the inside. We have here a construct, but so natural a one as to be practically unavoidable. An example of a construct involving a greater amount of invention is the stress in an elastic body. A stress is by definition a property of the interior points of a body which is connected mathematically in a simple way with the forces acting across the free surface of the body. A stress is then, by its very nature, forever beyond the reach of direct experience, and it is therefore a construct. The entire structure of a stress corresponds to nothing in direct experience; it is related to force, but is itself a six-fold magnitude, whereas a force is only three-fold

We have next to ask whether the stress, which we have invented to meet the situation in a body exposed to forces, is a good construct. In the first place, a stress has the same number of degrees of freedom as the observable phenomenon, for it is one of the propositions of the mathematical theory of elasticity that the boundary conditions, which are the experimental variables, uniquely determine the stress in a *given* body [*i.e.* a body of given elastic constants]; and of course it is at once obvious, by an inspection of the equations, that conversely a possible stress system uniquely determines the boundary conditions to the significant amount. There is, therefore, a unique one-to-one correspondence between a stress and the physical situation it was made to meet, and so far a stress is a good construct. Up to this point a stress, from the point of view of the operations in terms of

which it is defined, is a purely mathematical invention, which is justified because it is convenient in describing the behavior of bodies subjected to the action of force. But we wish now to go farther and ascribe physical reality to a stress, meaning by this that a stress in a solid body shall correspond to some real physical state of the interior points. Let us examine, from the point of view of operations, what the meaning of a statement like this may be. Since we now wish to ascribe an additional physical meaning to a stress beyond that of the mathematical operations in terms of which the stress was determined, there must exist additional physical operations corresponding to this meaning, or else our statement is meaningless. Now of course it is a matter of the most elementary experience that physical phenomena do exist which allow these other independent operations. A body under stress is also in a state of strain, which may be determined from the external deformations, or the strain at internal points may be made more vividly real by those optical effects of double refraction in transparent bodies which are now so extensively used in model experiments, or if the stress is pushed beyond a certain point, we have such new phenomena as permanent set or finally, rupture.

We have, then, every reason to be satisfied with our construct of stress. In the first place, from the formal point of view, it is a good construct because there is a unique correspondence between it and the physical data in terms of which it is defined; and in the second place we have a right to ascribe physical

reality to it because it is uniquely connected with other physical phenomena, *independent of those which entered its definition*. This last requirement, in fact, from the operational point of view, amounts to nothing more than a definition of what we mean by the reality of things not given directly by experience. Since now in addition to satisfying the formal requirements, experience shows that a stress is most useful in correlating phenomena, we are justified in giving to this construct of stress a prominent place among our concepts.

Consider now another construct, one of the most important of physics, that of the electric field. In the first place, an examination of the operations by which we determine the electric field at any point will show that it is a construct, in that it is not a direct datum of experience. To determine the electric field at a point, we place an exploring charge at the point, measure the force on it, and then calculate the ratio of the force to the charge. We then allow the exploring charge to become smaller and smaller, repeating our measurement of force on each smaller charge, and define the limit of the ratio of the force to the charge as the electric field intensity at the point in question, and the limiting direction of the force on a small charge as the direction of the field. We may extend this process to every point of space, and so obtain the concept of a field of force, by which every point of the space surrounding electric charges is tagged with the appropriate number and direction, the exploring charge having completely disappeared.



The field is, then, clearly a construct. Next, from the formal point of view of mathematics, it is a good construct, because there is a one to one correspondence between the electric field and the electric charges in terms of which it is defined, the field being uniquely determined by the charges, and conversely there being only one possible set of charges corresponding to a given field. Now nearly every physicist takes the next step, and ascribes physical reality to the electric field, in that he thinks that at every point of the field there is some real physical phenomenon taking place which is connected in a way not yet precisely determined with the number and direction which tag the point. At first this view most naturally involved as a corollary the existence of a medium, but lately it has become the fashion to say that the medium does not exist, and that only the field is real. The reality of the field is self-consciously inculcated in our elementary teaching, often with considerable difficulty for the student. This view is usually credited to Faraday, and is considered the most fundamental concept of all modern electrical theory. Yet in spite of this, I believe that a critical examination will show that the ascription of physical reality to the electric field is entirely without justification. I cannot find a single physical phenomenon or a single physical operation by which evidence of the existence of the field may be obtained independent of the operations which entered the definition. The only physical evidence we ever have of the existence of a field is obtained by going there with an electric charge and

observing the action on the charge [when the charges are inside atoms we may have optical phenomena], which is precisely the operation of the definition. It is then either meaningless to say that a field has physical reality, or we are guilty of adopting a definition of reality which is the crassest tautology.

There can be no question whatever of the tremendous importance of the concept of the electric field as a tool in thinking about, describing, correlating, and predicting the properties of electrical systems; electrical science is inconceivable without this or something equivalent. But in addition to this aspect of the field concept, the further tacit implication of physical reality is almost always present, and has had the greatest influence on the character of physical thought and experiment. Yet I do not believe that the additional implication of physical reality has justified itself by bringing to light a single positive result, or can offer more than the pragmatic plea of having stimulated a large number of experiments, all with persistently negative results. It is sufficient to mention the fate of the attempt of Faraday and Maxwell to ascribe a stress like that of ordinary matter to the ether, which failed because, among other reasons, nothing can exist in the ether analagous to the *strain* of ordinary matter, to indicate the unfruitfulness of the idea of physical reality. It seems to me that any pragmatic justification in postulating reality for the electric field has now been exhausted, and that we have reached a stage where we should attempt to

get closer to the actual facts by ridding the field concept of the implications of reality.

Another indispensable and most interesting construct is that of the atom. This is evidently a construct, because no one ever directly experienced an atom, and its existence is entirely inferential. The atom was invented to explain constant combining weights in chemistry. For a long time there was no other experimental evidence of its existence, and it remained a pure invention, without physical reality, useful in discussing a certain group of phenomena. It is one of the most fascinating things in physics to trace the accumulation of independent new physical information all pointing to the atom, until now we are as convinced of its physical reality as of our hands and feet.

A construct which had to be abandoned because it did not turn out to have physical reality, and which furthermore was not sufficiently useful in the light of newly discovered phenomena, was that of a caloric fluid.

The notion of "physical reality" is not of prime importance to this discussion of the character of our constructs; our definition of the meaning of physical reality may not appeal to everyone. The essential point is that our constructs fall into two classes: those to which no physical operations correspond other than those which enter the definition of the construct, and those which admit of other operations, or which could be defined in several alternative ways in terms

of physically distinct operations. This difference in the character of constructs may be expected to correspond to essential physical differences, and these physical differences are much too likely to be overlooked in the thinking of physicists. We must always be on our guard not to forget the physical differences between a thing like a stress in an elastic body and an electromagnetic field.

The moral of all this is that constructs are most useful and even unavoidable things, but that they may have great dangers, and that a careful critique may be necessary to avoid reading into them implications for which there is no warrant in experience, and which may most profoundly affect our physical outlook and course of action.

### THE RÔLE OF MATHEMATICS IN PHYSICS

Practically all the formulations of theoretical physics are made in mathematical terms; in fact to obtain such formulations is generally felt to be the goal of theoretical physics. It is then evidently pertinent to consider what the nature of the mathematics is to which we assign so prominent a rôle.

We have in the first place to understand why it is possible to express physical relations in mathematical language at all. I am not sure that there is much meaning in this question. It is the merest truism, evident at once to unsophisticated observation, that mathematics is a human invention. Furthermore, the mathematics in which the physicist is interested was developed for the explicit purpose of describing the

behavior of the external world, so that it is certainly no accident that there is a correspondence between mathematics and nature. The correspondence is not by any means perfect, however, but there is always in mathematics a precise quality to which none of our information about nature ever attains. The theorems of Euclid's geometry illustrate this in a preëminent degree. The statement that there is just one straight line between two points and that this is the shortest possible path between the points is entirely different in character from any information ever given by physical measurement, for all our measurements are subject to error. It is possible, nevertheless, to give a certain real physical meaning to the ideally precise statements of geometry, because it is a result of everyday experience that as we refine the accuracy of our physical measurements the quantitative statements of geometry are verified within an ever decreasing margin of error. From this arises that view of the nature of mathematics which apparently is most commonly held; namely that if we could eliminate the imperfections of our measurements, the relations of mathematics would be exactly verified. Abstract mathematical principles are supposed to be active in nature, controlling natural phenomena, as Pythagoras long ago tried to express with his harmony of the spheres and the mystic relations of numbers.

This idealized view of the connection of mathematics with nature could be maintained only during that historical period when the accuracy of physical measurement was low, and must now be abandoned.



For it is no longer true that the precise relations of Euclid's geometry may be indefinitely approximated to by increasing the refinements of the measuring process, but there are essential physical limitations to the very concepts of length, etc., which enter the geometrical formulations, set by the discrete structure of matter and of radiation. This is no academic matter, but touches the essence of the situation. There is no longer any basis for the idealization of mathematics, and for the view that our imperfect knowledge of nature is responsible for failure to find in nature the precise relations of mathematics. It is the mathematics made by us which is imperfect and not our knowledge of nature. [From the operational point of view it is meaningless to attempt to separate "nature" from "knowledge of nature".] The concepts of mathematics are inventions made by us in the attempt to describe nature. Now we shall repeatedly see that it is the most difficult thing in the world to invent concepts which exactly correspond to what we know about nature, and we apparently never achieve success. Mathematics is no exception; we doubtless come closer to the ideal here than anywhere else, but we have seen that even arithmetic does not completely reproduce the physical situation.

Mathematics appears to fail to correspond exactly to the physical situation in at least two respects. In the first place, there is the matter of errors of measurement in the range of ordinary experience. Now mathematics can deal with this situation, although somewhat clumsily, and only approximately, by

specifically supplementing its equations by statements about the limit of error, or replacing equations by inequalities—in short, the sort of thing done in every discussion of the propagation of error of measurement. In the second place, and much more important, mathematics does not recognize that as the physical range increases, the fundamental concepts become hazy, and eventually cease entirely to have physical meaning, and therefore must be replaced by other concepts which are operationally quite different. For instance, the equations of motion make no distinction between the motion of a star into our galaxy from external space, and the motion of an electron about the nucleus, although physically the meaning in terms of operations of the quantities in the equations is entirely different in the two cases. The structure of our mathematics is such that we are almost forced, whether we want to or not, to talk about the inside of an electron, although physically we cannot assign any meaning to such statements. As at present constructed, mathematics reminds one of the loquacious and not always coherent orator, who was said to be able to set his mouth going and go off and leave it. What we would like is some development of mathematics by which the equations could be made to cease to have meaning outside the range of numerical magnitude in which the physical concepts themselves have meaning. In other words, the problem is to make our equations correspond more closely to the physical experience back of them; it evidently needs some sort of new invention to accomplish this.

We return later, in discussing Lorentz's equations of electrodynamics, to the disadvantages arising from the present indiscriminating character of mathematics. In the meantime, we must recognize that there are very important advantages here, as well as disadvantages. All experience justifies the expectation that the laws of nature with which we are already familiar hold at least approximately and without violent change in the unexplored regions immediately beyond our present reach. By assuming an unlimited validity for the laws as we now know them, mathematics enables us to penetrate the twilight zone, and make predictions which may be later verified. It is only when we are carried too far afield that we must deprecate this characteristic of our mathematics.

There is another aspect of the use of mathematics in describing nature that is often lost sight of; namely, that any system of equations can contain only a very small part of the actual physical situation; there is behind the equations an enormous descriptive background through which the equations make connection with nature. This background includes a description of all the physical operations by which the data are obtained which enter the equations. For instance, when Einstein formulates the behavior of the universe in terms of the world lines of events, the events as they enter the equations are entirely colorless things, merely 3 space and 1 time coördinate. To make connection with experience there must be a descriptive background giving the physical contents of the events; for example, there may be the state-

ment that some of the events are light signals. This descriptive background is supposed to remain fixed, unaffected by any operations to which the equations themselves are subject. If, for example, the frame of reference of the equations is altered by changing its velocity, the physical significance of the descriptive background is supposed to remain unaltered, or rather no mention is usually made of this question at all. It would seem, however, that this matter needs some discussion. The descriptive background gets its meaning only in terms of certain physical operations. If the descriptive background remains unaltered when the uniform velocity of the frame of reference is changed, for instance, this means that the motion of the frame of reference does not at all affect the possibility of carrying out certain operations. This is pretty close to the restricted principle of relativity itself, which states that the form of natural laws is not affected by uniform velocity. Until a more careful analysis of the situation is made it would seem therefore that there is some ground for the suspicion that the principle of relativity is involved in the possibility of giving to physical phenomena a *complete* mathematical formulation, understanding "complete" to mean "including the descriptive background."

## CHAPTER III

### DETAILED CONSIDERATION OF VARIOUS CONCEPTS OF PHYSICS

WE now begin our detailed consideration of the most important concepts of physics. It is entirely beyond the scope of this essay to attempt more than an indication of some of the most important matters. Neither is it to be expected that the parts of this discussion will always have a very close connection with each other; the purpose of the discussion is to aid in acquiring the greatest possible self-consciousness of the whole structure of physics.

#### THE CONCEPT OF SPACE

A logically satisfying definition of what we understand by the concept of space is doubtless difficult to give, but we shall not be far from the mark if we think of it as the aggregate of all those concepts which have to do with position. Position means position of something. The position of things is determined by some system of measurement; perhaps the simplest is that implied in a Cartesian coördinate system with its three measurements of length. Hence much of the essential discussion of space has already been given in connection with the concept of length. We have seen that measurements of length are made with phys-



ical measuring rods applied to some physical object. We cannot measure the distance between two points in empty space, because if space were empty there would be nothing to identify the position of the ends of the measuring rod when we move it from one position to the next. We see, then, from the point of view of operations that the framework of Cartesian geometry, often imagined in an ideal mathematical sense, is really a physical framework, and that what we mean by spatial properties is nothing but the properties of this framework. When we say that space is Euclidean, we mean that the physical space of meter sticks is Euclidean: it is meaningless to ask whether empty space is Euclidean. Geometry, therefore, in so far as its results are expected to apply to the external physical world, and in as far as it is not a logical system built up from postulates, is an experimental science. This view is now well understood and accepted, but there was a time when it was not accepted, but vigorously attacked; the change of attitude toward this question is symptomatic of a change of attitude toward many other similar questions.

We have already emphasized that the space of astronomy is not a physical space of meter sticks, but is a space of light waves. We may, therefore, have different kinds of space, depending on the fundamental operations. The space of meter sticks we have called "tactual space", and the space of light beams "optical space". If we ask whether astronomical space is Euclidean, we mean merely to ask whether those features of optical space which are within the

reach of astronomical measurement are Euclidean. The only possible attitude with respect to this question, or such related questions as whether the total volume of space is finite, or whether space has curvature, is that it is entirely for experiment to decide, and that we have no right to form any preconceived notion whatever. It is therefore beyond the scope of this discussion.

It is interesting to notice that the restricted theory of relativity virtually assumes, although often without making the explicit statement, that tactual and optical space are the same. This equivalence results from the properties assumed for light beams. The distance of a mirror may be found equally well by measuring it with meter sticks, or by determining the time required by a light signal to travel there and back. This situation is, however, logically unsatisfying, because it must be assumed that the operations for measuring time are independently defined, and we shall see that they are not. It is a consequence of the assumed equivalence of tactual and optical space that the path of a beam of light is a straight line, as a straight line is determined by operations with meter sticks. When we come to astronomical phenomena, the physical operations with meter sticks can no longer be carried out, and it is meaningless to ascribe to beams of light on an astronomical scale the same geometrical properties that we do on a small scale.

### THE CONCEPT OF TIME

According to our viewpoint, the concept of time

is determined by the operations by which it is measured. We have to distinguish two sorts of time; the time of events taking place near each other in space, or local time, and the time of events taking place at considerably separated points in space, or extended time. As we now know, the concept of extended time is inextricably mixed up with that of space. This is not primarily a statement about nature at all, and might have been made simply by the observation that the operations by which extended time is measured involve those for measuring space. Of course historically the doctrine of relativity was responsible for the critical attitude which led to an examination of the operations of measuring time, but relativity was not necessary for a realization of the spatial implications of time, any more than the discovery of Planck's quantum unit  $h$  was necessary for the invention by Planck of his absolute units of measurement, although historically he was inspired to make this invention by discovering  $h$ , and in his own mind seems to have thought of the connection as a necessary one.<sup>1</sup>

The physical operations at the basis of the measurement of time have never been subjected to the critical examination which seems to be required. One method of measurement, for instance, involves the properties of light.

A meter stick is set up with mirrors at the two ends, and a light beam travels back and forth between

<sup>1</sup> Max Planck, *The Theory of Heat Radiation*, translated by Masius, P. Blakiston's Son & Co., 1914 edition, p. 174.

the two mirrors without absorption. The time required for a single passage back and forth is defined as the unit of time, and time is measured simply by counting these intervals. But such a procedure is unsatisfactory if we are to permit ourselves all those operations which are demanded by even the simplest postulate of relativity, for we must be able to move our clock from place to place, transfer it from one system to another in relative motion, and with it determine the properties of light beams in the stationary or moving system. We recognize in principle that the length of the meter stick may be different when it is in motion, that it may change also during the acceleration incident to moving it from one place to another, and that until proved to the contrary the velocity of light may be a function of velocity or acceleration. The complicated interplay of all these possibilities leaves us in much doubt as to the physical significance of such postulates as, for example, that the velocity of light is the same in the moving system and the stationary system. In order to ascribe any simple significance to postulates about the velocity of light, it would seem that we must have an instrument for measuring this velocity, and therefore for measuring time, which does not itself involve the properties of light. To do this we might seek to specify the measurement of time in purely mechanical terms, as for instance in terms of the vibration of a tuning fork, or the rotation of a flywheel. But here again we encounter great difficulties, because we recognize that the dimensions of

our mechanical clock may change when it is set in motion, and that the mass of its parts may also change. We want to use the clock as a physical instrument in determining the laws of mechanics, which of course are not determined until we can measure time, and we find that the laws of mechanics enter into the operation of the clock.

The dilemma which confronts us here is not an impossible one, and is in fact of the same nature as that which confronted the first physicist who had to discover simultaneously the approximate laws of mechanics and geometry with a string which stretched when he pulled it. We must first guess at what the laws are approximately, then design an experiment so that, in accordance with this guess, the effect of motion on some phenomenon is much greater than the expected effect on the clock, then from measurements with uncorrected clock time find an approximate expression for the effect of motion on mass or length, with which we correct the clock, and so on *ad infinitum*. However, so far as I know, the possibility of such a procedure has not been analyzed, and until the analysis is given, our complacency is troubled by a real disquietude, the intensity of which depends on the natural skepticism of our temperament.

In practice, the difficulties of such a logical treatment are so great that the matter has been entirely glossed over. It is convenient to postulate a clock, of unknown construction, but such that the velocity of light, when measured in terms of it, has certain prop-



erties. Such, for example, is the point of view in Birkhoff's recent book.<sup>1</sup> The difficulty with this method is that the resulting edifice is as divorced from physical reality as is the logical geometry of postulates. We cannot be at all sure that the properties of light as measured with our physical clocks are the same as the theoretical properties. The difficulty is particularly important and fundamental in the general theory of relativity; the basis of the whole theory is the infinitesimal interval  $ds$ , which is supposed to be given. Once given, the mathematics follows. But in a physical world,  $ds$  is *not* given, but must be found by physical operations, and these operations involve measurements of length and of time with clocks whose construction is not specified. In any actual physical application the question must be answered whether the physical instrument used in measuring the temporal part of  $ds$  is really a clock or not. There is at present no criterion by which this question can be answered. If the vibrating atom is a clock, then the light of the sun is shifted toward the infra-red, but how do we know that the atom is a clock (some say yes, others no)? If we find the displacement physically have we thereby proved that general relativity is physically true, or have we proved that the atom is a clock, or have we merely proved that there is a particular kind of connection between the atom and the rest of nature, leaving the possibility open that neither is the atom a clock nor

<sup>1</sup> G. D. Birkhoff. *Relativity and Modern Physics*, Harvard University Press, 1923.

general relativity true? In practice, of course, we shall adopt the solution which is simplest and most satisfying aesthetically, and doubtless shall say that the atom is a clock and relativity true. But if we adopt this simple view, we must also cultivate the abiding consciousness that at some time in the future troubles may have their origin here.

It seems to me that the logical position of general relativity theory is merely this: Given any physical system, then it is possible to assign values to  $ds$  such that relations mathematically deduced by the principle of relativity correspond to relations between measurable quantities in the physical system; but that the things that we physically call  $ds$  are anything more than approximately connected with the  $ds$ 's required to give the mathematical relations, is at present no more than a pious faith.

To return to the concept of time, we have already stated that there are two main problems, that of measuring time at a single point of space, and that of spreading a time system over all space. The second aspect of the problem is that to which attention has been directed by relativity theory; the following detailed examination shows how the operations of relativity for setting and synchronizing clocks at distant places involve the measurement of space. It is a fundamental postulate that the adjustment of the clocks is to be accomplished by light signals. The synchronization of the clocks is now simple enough. We merely demand that light signals sent from the master clock at intervals of one second arrive at any

distant clock at intervals of one second as measured by it, and we change the rate of the distant clock until it measures these intervals as one second. After its rate has been adjusted, the distant clock is to be so *set* that when a light signal is despatched from the master clock at its indicated zero of time the time of arrival recorded at the distant clock shall be such that the distance of the clock from the master clock divided by the time of arrival shall give the velocity of light, assumed already known. This operation involves a measurement of the distance of the distant clock, so that in spreading the time coördinates over space the measurement of space is involved by definition, and the measurement of time is, therefore, not a self-contained thing. This is the physical basis for the treatment of space and time as a four-dimensional manifold. Although mathematically the numbers measuring space and time enter the formulas symmetrically, nevertheless the physical operations by which these numbers are obtained are entirely distinct and never fuse, and I believe it can lead only to confusion to see in the possibility of a four dimensional treatment anything more than a purely formal matter.

The notion of extended time, therefore, involves the measurement of space. It is an interesting question whether the notion of local time also involves the measurement of space. A rigorous answer to this question involves giving the specifications for the construction of a clock, which we have seen has not yet been done. It seems to me probable, however, that the construction of even a single local clock

involves in some way the *measurement* of space. If, for example, we use a vibrating tuning fork, we must find how the time of vibration depends on the amplitude of vibration, and this involves space measurement, or if we use a rotating flywheel, we have to correct for the change of moment of inertia due to the change of dimensions when it is set into motion or brought into a gravitational field, and all this involves space measurement. However, these considerations are not certain, and perhaps the question is not important.

There is now the further consideration that actually in practice the concept of local time is not entirely divorced from that of extended time, for two bodies cannot occupy the same space at the same time, and the time of any event is actually measured on an instrument at some distance, communication being maintained by light or elastic signals. But experience convinces us that in the limit, as the phenomenon to be measured gets closer to the clock, there is no measurable difference, whether communication with the clock is maintained by light, or acoustical or tactual signals, so that we have come in physical practice to accept measurement of the time of events in the immediate neighborhood of the clock (local time) as one of the ultimately simple things behind which we do not attempt to go.

Local time is, therefore, a concept treated by the physicist even now as simple and unanalyzable. This concept is what most people have in mind when they think of time. Time, according to this concept, is

something with the properties of local time; it was something of this kind that Newton must have meant by his absolute time, and it is the tacit retention of this sort of concept that is responsible for the difficulty so often found in grasping the idea of the relativity of simultaneity, which is of course entirely foreign to our experience of simultaneity in local time. An examination of the operations involved in extending time has shown how the concept of extended time is different from that of simple local time; this difference leads to appreciably different numerical relations when we are dealing with high velocities or great distances. Local time is proved by experience not to be a satisfactory concept for dealing with events separated by great distances in space or with phenomena involving high velocities. For instance, we must not talk about the age of a beam of light, although the concept of age is one of the simplest derivatives of the concept of local time. Neither must we allow ourselves to think of events taking place in Arcturus *now* with all the connotations attached to events taking place *here now*. It is difficult to inhibit this habit of thought, but we must learn to do it. The naïve feeling is very strong that it does *mean* something to talk about the entire present state of the universe independent of the process by which news of the condition of distant parts is determined by us. I believe that an examination of this feeling will show that it is psychological in character; what we mean by the totality of the present is merely the entire present content of our conscious-



ness. This is apparently a simple direct thing; we do not appreciate until we make further analysis that our present consciousness of the existence of the moon or a star is due to light signals, and that therefore the apparently simple immediate consciousness of events distant in space involves complicated physical operations.

Similarly, if we continue to use local time, we get into trouble, when we go to high velocities, with our simple concept of velocity, which may be defined in terms of a combination of space and time concepts. The concept of local time thus loses its value and becomes merely a blunted tool when we try to carry it out of its original range. But the concept of extended time, with which we have to replace local time, is a complicated thing, to which we have not yet got ourselves accustomed; it may perhaps prove to be so complicated as never to be a very useful intuitive tool of thought.

All these considerations about time have been concerned only with intervals of such an order of magnitude that they are readily experienced by any individual. If we have to deal with intervals either very long or very short, it is obvious that our entire procedure changes, and consequently the concept changes. In extending the time concept to eras remote in the past, for example, we try as always, to choose the new operations so as to piece on continuously with those of ordinary experience. A precise analysis of the change in the concept of time when applied to the remote past does not seem to be of

great significance for our present physical purpose, and will not be attempted here. It is perhaps worth while to point out, however, that all our other concepts, as well as that of time, must be modified when applied to the remote past; an example is the concept of truth. It is amusing to try to discover what is the precise meaning in terms of operations of a statement like this: "It is true that Darius the Mede arose at 6:30 on the morning of his thirtieth birthday."

Of more concern for our physical purposes is the modification which the time concept undergoes when applied to very short intervals. What is the meaning, for example, in saying that an electron when colliding with a certain atom is brought to rest in  $10^{-18}$  seconds? Here I believe the situation is very similar to that with regard to short lengths. The nature of the physical operations changes entirely, and as before, comes to contain operations of an electrical and optical character. The immediate significance of  $10^{-18}$  is that of a number, which when substituted into the equations of optics, produces agreement with observed facts. Thus short intervals of time acquire meaning only in connection with the equations of electrodynamics, whose validity is doubtful and which can be tested only in terms of the space and time coördinates which enter them. Here is the same vicious circle that we found before. Once again we find that concepts fuse together on the limit of the experimentally attainable.

This discussion of the concept of time will doubtless be felt by some to be superficial in that it makes

no mention of the *properties* of the physical time to which the concept is designed to apply. For instance, we do not discuss the one dimensional flow of time, or the irrevocability of the past. Such a discussion, however, is beyond our present purpose, and would take us deeper than I feel competent to go, and perhaps beyond the verge of meaning itself. Our discussion here is from the point of view of operations: we assume the operations to be given, and do not attempt to ask why precisely these operations were chosen, or whether others might not be more suitable. Such properties of time as its irrevocability are implicitly contained in the operations themselves, and the physical essence of time is buried in that long physical experience that taught us what operations are adapted to describing and correlating nature. We may digress, however, to consider one question. It is quite common to talk about a reversal of the direction of flow of time. Particularly, for example, in discussing the equations of mechanics, it is shown that if the direction of flow of time is reversed, the whole history of the system is retraced. The statement is sometimes added that such a reversal is actually impossible, because it is one of the properties of physical time to flow always forward. If this last statement is subjected to an operational analysis, I believe that it will be found not to be a statement about nature at all, but merely a statement about operations. It is *meaningless* to talk about time moving backward: by definition, *forward* is the direction in which time flows.

## THE CAUSALITY CONCEPT

The causality concept is unquestionably one of the most fundamental, perhaps as fundamental as that of space and time, and therefore at least equally entitled to a first place in the discussion. But as ordinarily understood, there are certain spatial and temporal implications in the causality concept, so that it can best be discussed in this order after our examination of space and time.

There is an aspect of the causality concept that in many respects is closely related to the question of "explanation", for to find the causes of an event usually involves at the same time finding its explanation. But there are nevertheless sufficient differences to warrant a separate discussion.

It seems fairly evident that there was originally in the causality concept an animistic element much like that in the concept of force to be discussed later. The physical essence of the concept as we now have it, freed as much as possible from the animistic element, seems to be somewhat as follows. We assume in the first place an isolated system on which we can perform unlimited identical experiments, that is, the system may be started over again from a definite initial condition as often as desired.<sup>1</sup> We assume

<sup>1</sup> We must include in general in the concept of "initial condition" the past history of the system. In order not to make this condition so broad as to defeat itself, we have to add the observation that actually identity of past history is necessary over only a comparatively short interval of time. Logical precision seems unattainable here—the physical concepts themselves have not the necessary precision.

further that when so started, the system always runs through exactly the same sequence of events in all its parts. This contains the assumption that the course of events runs independent of the absolute time at which they occur—there is no change with time of the properties of the universe.<sup>1</sup> It is a result of experience that systems with these properties actually exist. An alternative way of stating our fundamental hypothesis is that two or more isolated similar systems started from the same initial condition run through the same future course of events. Upon the system given in this way, which by itself runs a definite course of events, we assume that we can superpose from the outside certain changes, which have no connection with the previous history of the system, and are completely arbitrary. Now of course in nature, as we observe it, there is no such thing as an arbitrary change, without connection with past history, so that strictly our assumption is a pure fiction. It is here

<sup>1</sup> As so often in physics, we appear to be doing two things at once here. It is doubtful whether we can give a meaning to "definite initial condition" apart from the future behavior of the system, so that we have no real right to infer from uniform future behavior both a constancy of the laws of nature, independent of time, and a constancy of initial condition. I very much question whether a thoroughgoing operational analysis would show that there are really two independent concepts here, and whether the use of two formally quite different concepts is anything more than a convenience in expression. It seems to me that it may be just as meaningless to ask whether the laws of nature are independent of time as it was to ask with Clifford whether the absolute scale of magnitude may not be changing as the solar system travels through space.



that the animistic element still seems to persist, although perhaps not necessarily. We regard our acts as not determined by the external world, so that changes produced in the external world by acts of our wills are, to a certain degree of approximation, arbitrary. The system, then, on which we are experimenting, is one capable of isolation from us in that we may regard ourselves as outside the system, and having no connection with it. The system, furthermore, is capable of isolation from the rest of the physical universe, in that events taking place outside the system have no connection with those taking place inside.<sup>1</sup> Experience gives the justification for assuming that physical isolation of this sort is possible. Actually, of course, isolation is never complete, but only partial, up to presumably any desired degree of approximation.

The statement that two exactly similar isolated systems, starting from the same initial conditions (including past history in the general idea of initial condition) will run through the same future course of events involves as a corollary that if differences develop in the behavior of two such apparently similar systems these differences are evidence of other previous differences. The thesis that this corresponds to experience may be called the thesis of essential connectivity and is perhaps the broadest we have: it

<sup>1</sup> Here again, the concept of "isolation" or "connection" is defined only in terms of the behavior of the system, and it is not clear whether this is really an operationally independent concept or not.

is the thesis that differences between the behavior of systems do not occur isolated but are associated with other differences. It is essentially the same thesis as that already mentioned in connection with "explanation", namely that it is possible to correlate any of the phenomena of nature with other phenomena.

If now the connectivity or correlation between phenomena is of a special kind, we have a causal connection; namely, if whenever we arbitrarily impress event A on a system we find that event B always occurs, whereas if we had not impressed A, B would not have occurred, then we say that A is the cause of B, and B the effect of A. By suitably choosing the event A, we may find the effect of any event of which the system is susceptible.

The relation between A and B is an unsymmetrical one, by the very nature of the definition, the cause being the arbitrary variable element, and the effect that which accompanies it. Furthermore, A may obviously be the cause of more than one event B, and may cause a whole train of events.

The causal concept analyzed in this way is not simple by any means. We do not have a simple event A causally connected with a simple event B, but the whole background of the system in which the events occur is included in the concept, and is a vital part of it. If the system, including its past history, were different, the nature of the relation between A and B might change entirely. The causality concept is therefore a relative one, in that it involves the whole system in which the events take place.

In practice we now take an exceedingly pregnant step and seek to extend the concept, and rid ourselves as much as possible of its relativity. It is a matter of experience that there are often a great number of systems in which A is the cause of B. In many cases the causal relation persists through such a very wide range of systems that we lose sight entirely of the system, and come to assume that we have an *absolute* causal connection between A and B. For instance, when I strike a bell, and hear the sound, the causal connection persists through such a great number of different kinds of system that I might think that here is an absolute causal connection. Such an absolute causal connection would mean that always under all circumstances, the striking of the bell is accompanied by a sound. But *all* conditions means only *all* those conditions covered by experiment. Thus in the case of the bell, all our experiments were made in the presence of the atmosphere. The causal connection between the striking of the bell and the sound should have been always recognized in principle as relative to the presence of the atmosphere. Indeed, later experiments in the absence of the atmosphere show that the atmosphere does play an essential part. Now as a matter of fact, the atmosphere is so comparatively easy to remove that we very readily include the atmosphere in the chain of causal connection. But if the atmosphere had been impossible to remove, like the old ether of space, our idea of the causal connections between the striking of the bell and its sound might have been quite dif-

ferent. In actual physical applications of the causality concept, the constant background which is maintained during all the variations by which the causal connection is established usually has to be inferred from the context.

It is a matter of perhaps universal experience that the event A is accompanied by not only one event, which is the effect of A by definition, but A entails a whole causal train of events. It seems to be a generalization from experience that the causally connected train of events started by A is a never ending train, provided the system is large enough. This is perhaps not necessary in the general case, but if the event A involves imparting external energy to the system, or the action of external force (momentum change), there can be no question.

That there is a causal train started by A is particularly evident if A and B are separated in space. Thus in the case of the bell, the impulse given to the air by the vibration of the bell is propagated through the air as an elastic wave, which thus constitutes the causal train of events. The phenomenon of propagation is characteristic of causal connections of a mechanical character, and is the justification for the introduction of the time concept in connection with the causality concept, where it now appears for the first time. It is evident that when a disturbance is propagated to a distant point, the effect *follows* the cause in time, as time is usually measured.

We extend this result, and usually think that the effect *necessarily* follows the cause. We now examine

whether this is a necessary result of the causality concept. If we are to talk about the time of events at different places, we must have some way of setting clocks all over space. If this is done arbitrarily, there is no necessary connection between the local clock times of a cause and its effect, but nevertheless the causality concept involves a certain temporal relation even in this most general case. Suppose that event A takes place at point 1 and its effect, event B, at point 2. We station a confederate at 2 who sends a light signal (or any other sort of signal) to 1, as soon as the event B occurs at 2. Then it is a consequence of the nature of the causality concept that the signal cannot arrive at 1 before event A occurs. For if it did arrive before A, we should merely omit to perform A, which by hypothesis is arbitrary, and entirely in our control, and then our assumption would be violated that the system is such that the event B occurs only when A also occurs. The same argument shows *a fortiori* that if the effect B occurs at the same place as its cause A, it cannot precede it in time. I cannot see that the nature of the causality concept imposes any further restriction on the time of B. The restricted principle of relativity, however, in postulating that no signal can be propagated faster than a light signal, virtually makes a further assumption about the temporal connection of causally connected events, namely, that the event B at 2 cannot occur before the arrival at 2 of a light signal which started from 1 at the instant that A occurred at 1. For if B did occur earlier, we could



use events A and B as a signaling code, thus violating our hypothesis.

There is thus a closest connection in time, when time is extended over space as the theory of relativity directs, between cause and effect, depending on their separation in space; from this arises the relativity concept of the causal cone, which in the four dimensional manifold of space-time divides the aggregate of all those events which may be causally related from the aggregate of those which are separated by such a small interval in time and such a large interval in space that communication by light signals and therefore a causal connection is not possible. Given now two events A and B which are related as cause to effect in one system of reference, then they must be causally related also in every other system of reference. For if they were not, we could by definition of causality suppress the event A in one of the systems in which the causal relation does not hold, and this, because of the nature of the concept of event, involves suppressing A in all the systems, thus violating our hypothesis of a causal connection in the original system. The concept of event involved in this argument will be examined later. It appears then, that the fundamental postulate of relativity (that the form of natural laws is the same in all reference systems) demands that the temporal order of events causally connected be the same in all reference systems.

The whole universe at this present moment is often supposed to be causally connected with all succeeding states. This means that if we could repeat experience,

starting from the same initial conditions, the future course of events would always be found to be the same. The truth of this conviction can never be tested by direct experiment, but it is something at which we arrive by the usual physical process of successive approximation. It is difficult to formulate precisely what we mean by "present" state of the universe, and there is every reason to think that such a formulation is not unique, but the concept contains the necessary implication that none of the events constituting the "present" can be causally connected. The events in distant places which constitute the present must be separated by an interval of time less than time required by light to travel between the two places.

The conviction, arising from experience, that the future is determined by the present and correspondingly the present by the past, is often phrased differently by saying that the present causally determines the future. This is in a certain sense a generalization of the causality concept. It is one of the principal jobs of physics to analyze this complex causal connection into components, representing as far as possible the future state of the system as the sum of independent trains of events started by each individual event of the present. How far such an analysis is possible must be decided by experiment. It is certainly possible to a very large extent in most cases, but there seems to be no reason to expect that a complete analysis is possible. So far as the system is describable in terms of linear differential equations, the

causal trains started by different events propagate themselves in space and time without interference and with simple addition of effects, and conversely the present may be analyzed back into the simple sum of elementary events in the past, but if the equations governing the motion of the system are not linear, effects are not additive, and such a causal analysis into elements is not possible. No emphasis is to be laid here on the *differential* aspect of the equations: it is quite possible that finite difference equations may have the same property of additivity. Although there can be no question that linear equations enormously preponderate, neither can there be any doubt that some phenomena cannot be described in terms of linear equations (*e.g.*, ferro-magnetism), so that there seems no reason to think that a causal analysis is always possible. I believe, however, that the assumption that such an analysis into small scale elements is possible is tacitly made in the thought of many physicists. If the analysis is not possible, we may expect to find results following the coöperation of several events which cannot be built up from the results of the events occurring individually.

When a causal analysis is possible, finding the simplest events which act as the origin of independent causal trains is equivalent to finding the ultimate elements in a scheme of explanation, so that here we merge with the concept of explanation, as already mentioned. As was true of the explanatory sequence, so here there can be no *formal* end of the causal sequence, because we can always ask for the cause

of the last member. But it may be physically meaningless to extend the causal sequence beyond a certain point. We have seen from the point of view of operations that the causal concept demands the possibility of variation in the system. It is therefore meaningless to say that A is the cause of B unless we can experience systems in which A does not occur. Now if in extending the causal sequence, we eventually arrive at a condition so broad that physically no further variation can be made, our causal sequence has to stop.

Corresponding to this property of the causality concept, the causal sequence may be terminated either formally, by postulate, or naturally, by the intrinsic physical nature of the elements of the sequence. Thus if we say that light gets from point to point because it is propagated by a medium of unalterable properties, which fills all space, which is always present and can never be eliminated physically, we have by the postulated properties of the medium brought the possibility of further inquiry to a close, because to take the next step and ask the cause of the properties of the ether, demands that we be able to perform experiments with the ether altered or absent. Such an ending of the sequence is evidently pure formalism, without physical significance. But other considerations may give physical significance. Thus if there are other sorts of experiment that can be explained by assuming a universal medium of the same properties, the concept proves not only to be useful, but to have a certain degree of physical sig-

nificance. An example of an inevitable termination of the causal sequence is afforded by the possibility, already mentioned, that the value of the gravitational constant may be determined by the total quantity of matter in the universe. Without further qualification, this is an entirely sterile statement, but if it can be shown that there is a simple numerical connection, the matter takes on interest, and we may seek further for a correlation between the numerical relation and other things.

This analysis of the causality concept does not pretend to be complete and leaves many interesting questions untouched. Perhaps one of the most interesting of these questions is whether we can separate into cause and effect two phenomena which *always* accompany each other, and whether therefore the classification of phenomena into causally connected groups is an exhaustive classification. But the discussion is broad enough for our purpose here; the most important points of view to acquire are that the causality concept is relative to the whole background of the system which contains the causally connected events, and that we must assume the possibility of an unlimited number of identical experiments, so that the causality concept applies only to sub-groups of events separated out from the aggregate of all events.

### THE CONCEPT OF IDENTITY

One of the most fundamental of all the concepts with which we describe the external world is that of identity; in fact, thinking would be almost inconceiv-



able without such a concept. By this concept we bridge the passage of time; it enables us to say that a particular object in our present experience is the same as an object of our past experience. From the point of view of operations, the meaning of identity is determined by the operations by which we make the judgment that this object is the *same* as that one of my past experience. In practice there are many indirect ways of making this judgment, but I believe the essence of the situation lies in the possibility of continuous connection between the object of the present and the past by continuous observation (either direct or indirect) through all intermediate time. We must, for example, be able to look continuously at the object, and state that while we look at it, it remains the same. This involves the possession by the object of certain characteristics—it must be a discrete thing, separated from its surroundings by physical discontinuities which persist. The concept of identifiability applies, therefore, only to certain classes of physical objects; no one thinks of trying to identify the wind of to-day with the wind of yesterday. It is somewhat easier to identify a liquid such as water in its flow in a stream, because we can make the motion of the water visible by solid particles suspended in it, but even here it is not easy to prove to a captious critic that it is really the water and not the suspended particles of solid that we are identifying. Even solids, when our measurements are sufficiently refined, seem to lose their discontinuous edges, as has been suggested in the discussion

of the approximate character of experimental arithmetic, and the identity concept becomes hazy.

There can be no question that the concept of identity is a tool perfectly well adapted to deal approximately with nature in the region of our ordinary experience, but we have to ask a more serious question. Does not the apparent demand of our thinking apparatus to be furnished with discrete and identifiable things to think about impose a very essential restriction on any picture of the physical universe which we are able to form? We are continually surprising ourselves in the invention of discrete structure further and further down in the scale of things, whose sole *raison d'être* is to be found entirely within ourselves. Thus Osborne Reynolds<sup>1</sup> has speculated seriously and most elaborately about an atomic structure in the ether, and we find Eddington<sup>2</sup> hinting at the existence of structure of an order of magnitude of  $10^{-40}$  cm. On a much larger scale of magnitude we also think in the same terms, and conceive positive and negative elementary charges with hard and impenetrable cores, which involves a complete change in the law of force at points sufficiently close. What physical assurance have we that an electron in jumping about in an atom preserves its identifiability in anything like the way that we suppose, or that the identity concept applies here at all? In fact, the iden-

<sup>1</sup> Osborne Reynolds, *The Sub-Mechanics of the Universe*, 254 pp., Cambridge University Press, 1903.

<sup>2</sup> A. S. Eddington, *Report on Gravitation*, *Lon. Phys. Soc.*, 1918, p. 91.

tity concept seems to lose all meaning in terms of actual operations on this level of experience.

The mind seems essentially incapable of dealing with continuity as a property of physical things; it is not even able to talk about continuity except in negative terms. To each attempted description of the properties of a truly continuous substance, it can say "No, it is not that", but cannot imagine experience which corresponds to what it conceives a really continuous thing ought to feel like. In terms of operations, continuity has only a sort of negative meaning. Now certain implications of this inability of the mind can be removed by appropriate postulates, as, for example, we can postulate the complete annihilation of a negative by a positive charge, as is now being done in certain speculations.<sup>1</sup> There is point in doing this, because the annihilation of two charges has physical meaning. But it is a question whether *all* implications of this habit of thought can be removed, and whether any picture that we can form of nature will not be tinged—sicklied o'er with the pale cast of thought.

The operational view suggests that in this last we are coming perilously close to a meaningless question, although there is a certain sense in which there is meaning here. It may turn out as a matter of fact that we shall not be able to carry our delving into small-scale phenomena deeper than a certain point, and that nature will appear to be finite downward, so that we shall bring up against a wall of some kind.

<sup>1</sup> For example, J. H. Jeans, *Nat.* 114, 828-829, 1924.

But to ask in such a situation whether we have come to the end because nature is *really* finite, or whether we only appear to be at an end because of some property of our minds, such as inability to deal with continuity, is, I believe, a meaningless question.

In actual use the identity concept is extended, and identity is used in other senses than the fundamental one examined above. For instance, we speak of two observers seeing the same object, or if the object moves or does something, we may speak of two observers perceiving the same happening. A happening about which the judgment of sameness is possible when perceived by different observers (or mathematically expressed when observed in two reference frames) is what we mean by an event, which is one of the fundamental concepts of relativity theory. What now is involved in this concept of event, or what do we mean when we say that two observers experience the same event? A first crude attempt might say that the event is the same if it is described in the same way by the two observers. But this leads us into all the complicated questions of the meaning of language, which we would gladly avoid, and is furthermore not true, because the whistle of a locomotive, for example, does not have the same pitch for two observers moving with different velocities. A satisfactory analysis of the situation is difficult to give, but I believe the essence lies in the discrete character of the event, just as the identity concept when applied only to objects involved discreteness. The event is bounded on all sides by dis-

continuities, both in space and time. Now it seems to be a result of experience that discontinuities have a certain absolute significance, in that there is a one-to-one correspondence between the discontinuities observed in any one reference system and those observed in any other. Corresponding discontinuities in two reference systems are by definition the same. An event is by definition the aggregate of all phenomena bounded by certain discontinuities, and two reference systems are by definition describing the same event if the discontinuous boundaries of the event are the same, irrespective of the appearance of the event in the two systems. The emission of a light signal, for example, is an event according to this definition, although it may appear as red light in one reference system and green in another.

We now see that the concept of event is only an approximate concept, as was also that of identity, and for the same reason, namely, there are no such things in experience as sharp discontinuities, but as our measurements become more refined, the edges of supposed discontinuities become blurred. As we go to smaller scales of magnitude this blurring becomes more important, until the physical possibility of performing those operations by which the discontinuities are detected entirely disappears, and the concept of event acquires, in terms of operations, an entirely different meaning. We continue to think of the event in the same way as before in terms of a mental model, but the true operational significance now depends on the particular phenomenon under consideration. The



concept of event is really not the same sort of thing when applied to the emission of a quantum of radiation from an atom, or the emission of gamma radiation from a radioactive disintegration, or the flashing of a signal from a dark lantern by opening and closing a shutter. Here as always, when our range of experience is extended, we must be prepared at some future time to find that, by extending the ordinary concept of event to small-scale phenomena by the device of the mental model, we have by implication smuggled into our picture phenomena which do not exist, so that it will be necessary to revise our thinking, casting it into terms corresponding to direct experience.

### THE CONCEPT OF VELOCITY

The concept of velocity, as ordinarily defined, involves the two concepts of space and time. The operations by which we measure the velocity of an object are these: we first observe the time at which the object is at one position, and then later observe the time at which it is at a second position, divide the distance between the two positions by the time interval, and if necessary, when the velocity is variable, take the limit. As long as we deal with fairly low velocities we do not have to inquire carefully as to the kind of time we use in these operations, but when the velocities become high, we do have to take care to use the local times at the two positions of the body, which means that we must have a time system spread over space, or, in other words, the "extended" time system. This velocity concept, defined in this way,

may be used as a tool in describing nature, and it will be found that nature has certain properties; for example, the velocity of light is  $3 \times 10^{10}$  cm./sec. Further, no material thing can be given a velocity as high as this, but as its velocity is made to approach this value, increments of energy increasing without limit are required.

But now it is very much a question for examination whether the velocity concept defined in this particular way has been chosen wisely as a tool for describing natural phenomena. It is quite possible to modify the velocity concept, that is, to set up other operations which correspond to our instinctive feeling of what velocity is in terms of immediate sensation and such that all numerical measures are unmodified at low velocities.<sup>1</sup> For example, a traveller in an automobile measures his velocity by observing the clock on his instrument board and the mile stones which he passes on the road. This operation differs from that of the definition above in that the time is no longer extended time, but is the local time of the moving object. The space coördinates used in this alternative operation at first seem a hybrid sort of thing, but they are what the observer would actually most naturally use: they are what he would measure with a tape measure fixed to a point of the road and allowed to

<sup>1</sup> It is an interesting question for the psychologist whether the velocity concept is not a more primitive thing in order of apprehension than that of time, and whether the concept of time is not derived from observing things in motion, or whether indeed there is any necessary connection at all between velocity and time in terms of untutored experience.

unwind as he proceeds, or what is measured by a vessel at sea with a log line let out behind. Or there is still another most interesting way of defining velocity, in which the analysis into space and time is not made at all, but velocity is directly measured by building up the given velocity by physical addition of a unit velocity selected arbitrarily. This matter is discussed at some length in my book "*Dimensional Analysis*",<sup>1</sup> but is of sufficient pertinence here to describe briefly. We may in the first place construct a concrete standard for velocity, as, for example, by stretching a string between two pegs on a board with a fixed weight. If we strike the string, a disturbance travels along the string which we can follow with the eye, and we define unit velocity as the velocity of this disturbance. An object has greater than unit velocity if it outruns the disturbance, and less if it lags behind. We may now duplicate our standard, making another board with pegs and stretched string, and check the equality of the two velocities by observing that the two disturbances run together. We now define two units of velocity as the velocity of anything which runs with the disturbance of the string of the second board, when the second board is made to move bodily with such a velocity that it runs with the disturbance of the first string. The process may be extended indefinitely, and any velocity measured.

If either of these two alternative definitions of velocity were adopted, it would be found that the

<sup>1</sup> Yale University Press, 1922.

velocity of light is infinite. Further, there would be no limit to the velocity which can be imparted to material bodies on giving them unlimited energy, which is what we are prepared to regard from ordinary experience is natural and simple. The infinite velocity for light, on the other hand, is most unnatural, particularly if we favor a medium point of view. We are here faced with a dilemma—all sorts of phenomena cannot at the same time be treated simply. If we attach the most fundamental significance to the behavior of material bodies, we shall do well to adopt one of the alternative definitions of velocity. If, on the other hand, we regard the phenomena of light as the most fundamental, we shall endeavor to form our definition so that the properties of light are simple. This was precisely the point of view of Einstein; it is characteristic of his entire scheme of restricted relativity that light is the fundamental thing, and this influenced him in adopting the first definition of velocity. Now one can have no quarrel with this desire to make light fundamental (the wisdom of doing this is to be justified by the results), and if the properties of light are to be treated mathematically, one can easily see the desirability of getting rid of infinite attributes, and so admit the desirability of making the velocity of light finite. But all this involves another very insidious assumption which we ourselves have tacitly used in all our preceding discussion, namely, that the notion of velocity properly pertains to light at all. Einstein has very definitely adopted this point of view, and

so determined the character of the entire structure of relativity. I believe, on the contrary, that it is very gravely to be questioned whether the identification of light with a *thing travelling*, which is involved in applying the velocity concept, should be made. This discussion must be postponed, however, until we deal with the properties of light. The important points for us to notice at present are that the definition of velocity actually used involves the concept of extended time, and that it would be possible to define velocity in different ways, which would give quite a different complexion to phenomena at high velocities, but which would leave untouched our ordinary experience.

The velocities at which the precise form of definition becomes important are higher than can be reached in ordinary mechanical experiments. Such velocities can be attained in terrestrial laboratories only with electrified particles, as in experiments in high vacua or with radioactive disintegrations. It is interesting to notice that we very seldom attempt a direct measurement of velocity in such experiments by following a discrete particle in its flight and finding the time required to pass over a measured distance, but the velocities are measured indirectly, by calculation from the equations of electrodynamics and in terms of such immediately observed things as curvature of path. It is true that one or two experiments have attempted a more direct measure of velocity, but it seems there is room for more work here.



## THE CONCEPTS OF FORCE AND MASS

Another concept of great importance is that of force. Since the usual analysis finds a connection between force and acceleration, and acceleration involves velocity, this is a natural place for the discussion of force. This concept has been subjected to much analysis by various writers. In origin the concept doubtless arises from the muscular sensations of resistance experienced from external bodies. This crude concept may at once be put on a quantitative basis by substituting a spring balance for our muscles, or instead of the spring balance we may use any elastic body, and measure the force exerted by it in terms of its deformation. Of course, the various precautions which must be taken in carrying out this idea physically are complex; the matter of precautions against temperature changes, for example, is one of the most easily understood. The concept of force so defined is limited to static systems; it is the task of statics to find the relation between the forces in systems at rest. We next extend the force concept to systems not in equilibrium, in which there are accelerations, and we must conceive that at first all our experiments are made in an isolated laboratory far out in empty space, where there is no gravitational field. We here encounter a new concept, that of mass, which as it is originally met is entangled with the force concept, but may later be disentangled by a process of successive approximations. The details of the various steps in the process of approximation are very instructive as typical of all methods in physics,

but need not be elaborated here. Suffice it to say that we are eventually able to give to each rigid material body a numerical tag characteristic of the body, such that the product of this number and the acceleration it receives under the action of any given force applied to it by a spring balance is numerically equal to the force, the force being defined, except for a correction, in terms of the deformation of the balance, exactly as it was in the static case. In particular, the relation found between mass, force, and acceleration applies to the spring balance itself by which the force is applied, so that a correction has to be applied for a diminution of the force exerted by the balance arising from its own acceleration.

We now extend the scope of our measurements by bringing our laboratory into the gravitational field of the earth, and immediately our experience is extended, in that we continually see bodies accelerated with no spring balance (that is, no force) acting on them. We extend the concept of force, and say that any body accelerated is acted on by a force, and the magnitude of this force is defined as that which would have been necessary to produce in the same body the same acceleration with a spring balance in empty space. There is physical justification for this extension in that we find we can remove the acceleration which a body acquires in a gravitational field by exerting on it with a spring balance a force of exactly the specified amount in the opposite direction. This extended idea of force may also be applied to systems in which there are electrical actions.

We thus see that in extending the notion of force from bodies in rest to bodies in motion, the character of the concept has changed, because the operations by which force is measured change—the force acting on a body is now measured in terms of its acceleration. But in determining the force from the acceleration, we have to know the mass. This mass has to be independently measured with the original concept of force; otherwise we have no basis for such simple statements as that the force of gravity on a body is proportional to its mass. All this applies to the ordinary range of experiments with low velocities. If now we extend the range of measurements, we find phenomena which we had not expected; for example, there seem to be difficulties in the way of indefinitely increasing the velocity of a material body, as of a charged atom. We begin to ask searching questions: is the force of gravity independent of velocity at high velocities, or is the mass independent of velocity under the same conditions or independent of the gravitational field, etc.?

In attempting to answer these new questions, we find difficulty with the concepts in terms of which they are formulated. There are no operations by which we can find whether force is independent of velocity unless we first know the mass, or any operations by which a mass can be measured unless we know a force. The purely mechanical systems with the highest velocities of which we have any experimental knowledge are the heavenly bodies. The motion of these is, with the important exception of

Mercury, that predicted by the ordinary laws of mechanics, so that at first it might appear that we have here confirmation of the laws of mechanics for bodies with comparatively high velocities. But it must be remembered that all we can observe of the heavenly bodies is their positions, and that we cannot perform on these bodies all the operations by which we can check the laws of mechanics for terrestrial phenomena. If, for example, mass and the force with which gravity acts on mass were both equally affected by velocity, the motion of the heavenly bodies would be exactly the same as that observed now. Hence as we increase the range of velocity, the concepts of force and mass simultaneously lose their definiteness, and become partially fused. This is typical of what we have now come always to expect near the limit of the experimentally attainable; experience becomes less rich, the choice of physical operations more restricted, concepts change and become fewer in number. If we are to retain the same formal number of concepts, we must introduce arbitrary conventions or definitions. These definitions are to be determined largely by convenience. In the case of mechanical systems, this motive of convenience is supplied by considerations from outside the domain of mechanical phenomena. The highest velocities of practice are not reached in mechanical, but in electrical systems, in experiments with vacuum tubes, etc. Considerations of convenience are therefore dictated from the electrical point of view. These considerations will be gone into in

much more detail later; the conclusion is all that we need here, which is that it is convenient to assume for the charge of the electron a constant number, independent of the velocity, and this involves making its mass variable in a definite way with velocity. Now if the principle of relativity is accepted, the mass of mechanical objects must vary with velocity in the same way as the mass of electrical charges. Since the variability of this latter is fixed, mechanical mass becomes a definite function of velocity, and the force is therefore also fixed in any specific physical case.

The fundamental definition of force given above is highly academic, involving as it does hypothetical experiments in laboratories situated far out in empty space. Some sort of procedure like this seems to correspond to more or less explicit statements to be found in the literature of mechanics. The meaning in terms of actual operations to be given to such definitions involves complicated inferential reasoning. We would make much closer connection with the conditions of actual experiment if in the definition we substituted for the hypothetical operations in empty space more or less approximately realizable operations on bodies sliding on level table tops without friction. I suppose our instinctive feeling for the laws of mechanics is such that we are convinced that definitions in terms of an interstellar space laboratory or a level table top are actually the same. But in principle we must recognize that when the operations are different, the concepts are different, and if we adopt something equivalent to the table



top definition, as it seems we are physically forced to do, we must leave open in our thinking the possibility of finding in the present penumbra, when our accuracy is sufficiently increased, such phenomena perhaps as directional attributes of mass in a gravitational field.

We have just considered the sort of problem that we encounter on ordinary scales of magnitude on going from low to high velocities; what becomes of the concepts of force and mass when we go to a very small scale? Down to the atomic scale we may at least slur over the new physical difficulties, for although we cannot of course experiment with actual atoms, we can nevertheless make measurements of the Brownian<sup>1</sup> movement of suspensions in liquids settling in a gravitational field, for example, and the extrapolation to the atom is not a very great one. The mass of each individual atom is obtained by what is equivalent to a process of counting, assuming the law of conservation of mass on an atomic scale. This is justified by all chemical experience. The mass of the component parts of the atoms, the electrons, may also perhaps be given a unique significance after we have decided on the laws of the electrical field, by experiments on acceleration in electrical fields. The question which interests in principle here is what meaning, if any, shall be attached to the mass of the elements of the electron.

<sup>1</sup> This phenomenon is discussed at length in the book by J. Perrin, *Brownian Movement and Molecular Reality*, translated by F. Soddy, Taylor and Francis, London, 1909.

It is evident that we here go beyond any possible experience, at least for the present, and that experience has again become poorer and our concepts fewer in number. All that we can now demand is that certain combinations of numbers, some of which represent mechanical mass and others electrical charge, have proper relations to each other when integrated throughout the entire body of the electron. Similar questions confront us when we ask what are the forces which the parts of the electron exert on each other. We return to this question in considering the nature of the electrical concepts. In any event, the concepts of both force and mass are entirely altered in this domain.

It is interesting to note, in passing, that present electrical theory gives no meaning to the mass of the elements of the electron, since the total electromagnetic mass of the electron is built up from the *mutual* terms in the action of the elements—the total mass is not a *linear* resultant of the action of the elements.

### THE CONCEPT OF ENERGY

In examining the concept of energy, we start with purely mechanical energy. In isolated mechanical systems, in which there are only conservative forces, the sum of kinetic and potential energy is constant. The kinetic energy may be defined as  $\Sigma \frac{1}{2} mv^2$ , formed for all parts of the body. The potential energy is determined by the position of the parts of the system, and has physical significance only with

reference to a datum position, that is, only changes of potential energy have meaning in terms of operations. The total energy ascribed to the system has therefore an element of arbitrariness in that the datum position may be chosen at random, and energy acquires meaning only on tracing the history back to the epoch of the datum position.

The concept of energy may be extended from mechanical systems to all systems with which we are acquainted; the operations by which meaning is given to the extended energy concept involve the generalized conservation principle, or the first law of thermodynamics. The extension to thermal systems is immediate; the inclusion of optical and electrical systems in the scheme was a most important physical step, which of course required careful experimental justification. Because of its wide range of application, the energy concept has now come to be regarded as one of the most important in physics; this idea was held by Ostwald<sup>1</sup> twenty and more years ago, and is now much to the front because of the connection between mass and energy indicated by the theory of relativity, and the important rôle assigned to energy levels in spectrum analysis.

What now is the precise nature and significance of the general energy concept? In the first place the conservation property of energy is one of the simplest and most obvious of the properties of matter, so that in this property of energy is seen a reason for ascribing to it certain of the properties of matter, in

<sup>1</sup> W. Ostwald, *Die Energie*, Barth, Leipzig, 1908.

particular and most important, that of localization in space. We must recognize, however, that this idea of a location in space is injected into the situation entirely by ourselves, and corresponds to nothing directly given by the operations of experiment. The idea has had a most important effect, however. Witness, for instance, the importance ascribed to the discovery by Kelvin of a function by which the total energy of an electric field can be represented as distributed through space;<sup>1</sup> this was one of the most important props of the medium point of view.

A more critical examination is likely to diminish considerably our satisfaction with this naïve analogy drawn between matter and energy. With regard to matter, we may still be tolerably satisfied with our ascription to matter of location in space, but it is quite different with regard to conservation of matter. In just what sense is matter conserved? Certainly not in terms of mass, as we at one time thought. Nevertheless we undeniably have a feeling that there is some sort of conservation property here, and are driven to formulate it badly in terms of a hypothetically constant number of protons and electrons. I have long thought that Newton was groping after some very similar idea when he so far forgot himself as to define mass as quantity of matter, a definition

<sup>1</sup> This function is  $\frac{1}{8}\pi$  times the scalar product of electric force and displacement. If Maxwell's definition of displacement is adopted, the factor  $\frac{1}{8}\pi$  is replaced by  $\frac{1}{2}$ , and an accurate analogy results between the energy stored in the ether and the elastic energy stored in a bent spring.

perfectly meaningless to a rigorous and unsympathetic interpretation. On the other hand, whatever meaning may reside in our idea of conservation of matter, it certainly is not, in at least one important respect, like the conservation of energy. For the energy of an isolated mechanical system is a function of the frame of reference in which it is described; merely by giving velocity to the reference frame and altering in no way the mechanical system we may change its kinetic, and so its total, energy by any direct amount. This does not even remotely resemble ordinary matter. I cannot see that the operations which are equivalent to the energy concept justify us in saying more than that energy is a property of a material system; the operations do not seem to give any unique meaning to a location associated with energy.

We now ask what significance is to be ascribed to the sort of conservation that energy does have. We restrict ourselves first to mechanical systems. The motions of a mechanical system satisfy certain differential equations of the second order, and the actual motion is to be found by an integration of the equations. In the integral of a differential equation certain constants appear which are determined by the initial conditions, and are therefore the same during all the future motion of the system; obviously these constants of motion correspond to conservative properties. This reasoning can of course be at once extended. Any system, mechanical or not, whose motion is determined by differential equations, will



have certain conservative properties. For the systems of mechanics energy is one of the conservative functions; others are momentum and moment of momentum. Energy is particularly simple, in that it is connected with measurable properties of the system by a simple formula ( $\Sigma \frac{1}{2} mv^2$ ), and is furthermore scalar, which is also a property of quantity of matter. But to go further and ascribe to energy other properties of matter, such as localization in space, is entirely overlooking the essential difference in the character of the operation by which matter and quantity of energy are measured, that is, overlooking the essential difference in their physical character.

The possible extension of the energy concept from mechanics to thermodynamics receives a sufficient physical explanation in terms of our views of the essentially mechanical character of thermal phenomena. That the idea can be extended also to simple electrical or magnetic systems, in which the effect of velocity of propagation is neglected, is a consequence of the fact that in these systems the equations of motion remain of the same general mechanical type, it having been shown by Maxwell that the equations of such systems may be written in the generalized Lagrangean form. When, however, we extend our formulas to systems in which the velocity of propagation is important (that is, when we consider the field equations in their general form) we find that the Lagrangean equations no longer apply to matter taken by itself, and energy is no longer conserved in the original sense. A new function appears, however,

which behaves mathematically in the same way that the energy did before. The equations of motion of the system remain Lagrangean in form if the mechanical parts of the system are supplemented by the electric and magnetic fields in space. In this extended form we have, therefore, a conservative function as before, and the energy concept may be retained in this enlarged aspect. The physical operations by which energy is determined are entirely altered, however, and the physical character of the concept is changed. No more than before is there justification for localizing energy in space, or ascribing to it other properties of matter. Yet the materialization of the energy concept, and the consequent desire that energy be localized in space, is one of the strongest arguments in many minds for the existence of a medium.

As far as I can see, therefore, the existence of conservative functions is involved in the possibility of describing natural phenomena with differential equations. That further there is a conservative function of the precise form found in mechanics is a consequence of the particular form of the equations and the nature of the forces. The question of the significance of the fact that the forces of nature appear to be conservative, with respect to this particular function of mechanics, is of much interest, but it is not our immediate concern now. We are interested rather to ask under what general conditions we shall have conservative functions. Quantum theory strongly suggests that when we pass to phenomena on a small enough scale, we may no longer be able to

employ differential equations in our descriptions, and hence the previous reason for the existence of constants connected with the motion disappears. Now there is one obvious remark to be made about this more general situation. Whenever the future history of a system is so connected with its present condition that we can retrace our way to the present from any future configuration, we shall always have conservative functions. For any future configuration contains certain fixed (or conservative) features, in that we can reconstruct the unique present from any future state. There is no reason to expect that the operations by which we find the fixed features will always be simple, as in the mechanical case. Now the determination of the future by the present, and conversely the possibility of reconstructing the present from the future (or the past from the present), is, we are convinced, a property which is at least approximately true of phenomena down to a smaller scale of magnitude than we have yet reached, and so we expect to find these conservative functions in systems whose ultimate laws of motion are much more general than any with which we are yet familiar. The particular form of the conservative function depends on the character of the system. That there is a scalar conservative function for ordinary systems depends of course on particular properties of the system, but we are at least prepared to find that a scalar conservative function does not necessarily mean a differential equation of the second order.

The potential energy of a system has a particular

significance with respect to this point of view. In an ordinary mechanical system, the potential energy simply measures the work done by the applied forces in being displaced from the initial to the final positions; that is, the potential energy is a measure of the deviation from the initial position, and so measures a certain feature of the history of the system. In the more general system, in which we may not have differential equations, we may look for something analogous to the potential energy which shall measure the displacement of the system from its initial configuration. Such a measure is always possible as long as the past can be reconstructed from the present (or the present from the future). We recall a remark of Poincaré's <sup>1</sup> to the effect that we of necessity must always have conservation, because if we have a system in which conservation apparently fails, we merely have to invent a new form of potential energy. This remark is obviously not of entire generality, but applies only to such systems as those considered here, in which the past may be reconstructed from the present.

Of late there has been much discussion of the advisability, on the basis of certain quantum phenomena, of giving up conservation as a principle applied to the details of the emission and absorption of light, retaining it only in a statistical sense. It seems to me that the question here in the minds of

<sup>1</sup> Henri Poincaré, *Wissenschaft und Hypothese*, translated into German by F. and L. Lindemann, Teubner, Leipzig, 1906, Chap. VIII.

physicists was always merely one of convenience, and that few, if any, doubted the ultimate applicability of the principle of Poincaré, or thought that we were here concerned with a system of such great generality that the past could not be reconstructed from the present. The question was merely whether those variables in terms of which the potential energy is defined make close enough connection with other things of immediate experimental significance, or whether on the whole the retention of a potential energy is not more trouble than is justified by its convenience, making a treatment from the statistical point of view preferable. However, this is all now a matter of more or less past history, because with the recent extension of the experiments of Compton,<sup>1</sup> we seem to have experimental evidence for the validity of the conservation law in detail for elementary quantum processes, with a corresponding simple potential energy.

Going still deeper, however, there are quantum phenomena which still may have to be treated by statistical methods, and this may mean giving up conservation in detail. We have no experimental evidence, for example, of what the electron is doing while jumping from one quantum orbit to another. A situation like this merely means that those details which determine the future in terms of the past may lie so deep in the structure that at present we have no immediate experimental knowledge of them, and we

<sup>1</sup> W. Bothe and H. Geiger, *ZS. f. Phys.* 32, 639-663, 1925. A. H. Compton, *Proc. Nat. Acad. Soc.*, 11, 303-306, 1925.



may for the present be compelled to give a treatment from a statistical point of view based on considerations of probability. But I suppose that no one, except perhaps Norman Campbell,<sup>1</sup> will maintain that such a situation is more than temporary, or will cease to search for consequences of these details of structure which may be open to experimental verification.

Similarly, we cannot permanently be satisfied with a picture of radioactive phenomena which represents radioactive disintegration as a matter of chance.

The general conclusion to which all this discussion leads is that energy is probably not entitled to the fundamental position that physical thought is inclined to give it, but that it is a more or less incidental consequence of more deep-seated properties, and that the character of these deep-seated properties is subject to only the most general restrictions, so that very little about the nature of the details can be inferred from the existence of any energy function.

### THE CONCEPTS OF THERMODYNAMICS

We shall not be concerned here with the many technical questions which are the proper subject of treatises on thermodynamics, but shall attempt an examination only of some fundamental concepts.

The most fundamental of these, which sets thermodynamics off apart from the simpler subjects, is probably that of temperature. In origin this concept was

<sup>1</sup> Norman Campbell, *Time and Chance*, *Phil. Mag.* 1, 1106-1117, 1926.

without question physiological, in much the same way as the mechanical concept of force was physiological. But just as the force concept was made more precise, so the temperature concept may be more or less divorced from its crude significance in terms of immediate sensation and be given a more precise meaning. This precision may be obtained through the notion of equilibrium states. We have in the first place the fundamental experimental fact that when a small body is placed inside a large system, which we recognize by crude means as comparatively invariable in temperature as time goes on, the small body very soon acquires a steady condition, that is, it comes to equilibrium with its surroundings. We now have the further experimental fact that if the small body A is in equilibrium with its environment, and body B is also in equilibrium with the same environment, there will be no change of condition of A and B when they are brought into contact with each other—that is, A and B are each in equilibrium with the other and also with the environment and therefore, by definition, at the same temperature as the environment. The temperature of the environment is now measured in terms of some of the properties of A and B which crude experience has shown change with the physiological temperature of A and B. The physiological notion of temperature is thus made more precise by being connected with the physical phenomenon of equilibrium.

Now it is at once evident that stated in this way without qualification we have said things that are not

true. It is not true in general that, when A is in equilibrium with an environment and B is in equilibrium with the same environment, A will be in equilibrium with B. Suppose, for example, that the environment is a stream of water and A is a tiny water wheel moving freely in its bearings, and that B is a similar wheel with much friction. Then we know that B will become warm, and will not be in equilibrium with A when brought into contact with it. Or we may choose for A a mercury thermometer with bulb covered with putty, and for B a similar thermometer with bulb sheathed in platinum, and we know that the two thermometers will not register the same temperature in the water stream. Or still more simply, we may try to read the temperature of the air in our garden on any bright day with a silvered and with a blackened bulb thermometer; we know that the two thermometers will read different temperatures. It is evident, therefore, that we shall have to specify much more carefully the conditions under which equilibrium holds if we are to give precise significance to the temperature concept.

It seems fairly evident in the first place that we shall have to rule out systems in which there is large scale mechanical motion; the simple notion of temperature does not apply to a system moving with respect to us. Only when the two thermometers A and B move with the same velocity as the stream do we have three-fold equilibrium between the stream, A and B. We may state this in another way by saying that the temperature of a moving body must be

measured on a thermometer stationary with respect to the body, but this is only a matter of words, and properly speaking the temperature concept applies only to a certain aspect of the relation between two bodies mutually at rest. We here entirely neglect relativity questions such, for example, as the proper way of correcting for the change of dimensions of moving thermometers.

If now the body whose temperature we are measuring does not move with the same velocity in all its parts, we may still give a meaning to local temperature by dividing the body into parts so small that the velocity of each part is essentially uniform, and measuring the temperature of each part with a thermometer stationary with respect to it. We are now confronted with the question of how far to carry the process of subdivision. Suppose we have a fluid whose motion is completely turbulent when measured with instruments of the ordinary scale of magnitude. For such a fluid the fundamental equilibrium proportions hold between two measuring bodies A and B and the fluid, provided that the bodies A and B are so large that the motion is completely turbulent on their scale of magnitude. We may then define the temperature of the turbulent fluid from the standpoint of these large scale bodies. But we may also define the temperature from the small scale point of view as the average of the temperatures recorded by sufficiently small thermometers, each moving with the velocity of a local bit of the fluid. These two temperatures will in general be different, and we

must more or less arbitrarily select one which we define as the true temperature. It would seem that the small scale temperature is the better one to choose, because there is a certain degree of arbitrariness in specifying the scale from which the motion shall be judged completely turbulent. But on the other hand, there are difficulties in the small scale definition, because the turbulence may become more and more fine grained, until we end with the motion of the molecules themselves, when the operations certainly fail which give meaning to the temperature concept. In this case of molecular turbulence, we are driven back to the large scale definition, which obviously corresponds to ordinary physical practice.

It appears then that the temperature concept is not a clean cut thing, which can be made to apply to all experience, but that it is more or less arbitrary, involving the scale of our measuring instruments. In any special case, the meaning of the temperature concept must be set by special convention. In practice this does not often make difficulty, because in the majority of cases there is no large scale motion with respect to the thermometers.

Consider now the other aspect of the equilibrium problem suggested by the thermometers with blackened and silvered bulbs in the sunshine. Our common experience tells us how to deal with this situation effectively enough for ordinary purposes. We recognize that the possibility of temperature equilibrium is disturbed by the radiation, and we protect the bulbs



of the thermometers from the sun's radiation by appropriate shields. But this only minimizes the difficulty. For the shield is warmed by the sun, and in turn warms to a less degree by its radiation the bulb of the thermometer within. We must recognize that every body, no matter what its temperature, is always emitting radiation, so that the bulb of our thermometer is always in a radiation field. At first this puts us in a serious quandary as to the whole question of equilibrium and the meaning of temperature. The situation is saved by the experimental observation that there is a particular radiation field which affects all thermometers equally, namely, the field inside an infinite body all at the same temperature. Logically this looks like the vicious circle again, for we have not yet defined what we mean by the same temperature. But actually we avoid the circle here, as in so many other physical cases, by a process of asymptotic approximation. The procedure is perhaps something like this: we find that if we experiment with larger and larger bodies, isolated and at great distances from other bodies at approximately the same temperature as judged by crude physiological sensations, two thermometers, identical except that the bulb of one is blackened and that of the other is silvered, record more nearly the same temperature as time goes on and as the thermometers are sunk to greater depths in the body. In actual practice, of course, the radiational opacity of most bodies is so high that these precautions against the effects of external radiation can usually be entirely ignored.

At high temperatures, on the other hand, radiation has to be explicitly dealt with.

The conclusion for us from these considerations is that operationally the concept of temperature is tied up with that of radiation—the equilibrium concept of temperature is strictly never exactly applicable; it is only a limiting sort of concept applicable when the radiation field is of a special sort, namely, that of a black body.

In spite of the explicit recognition which we have to give radiation in defining temperature, we usually entirely lose sight of it in thinking about the mechanism of ordinary physical processes, as for instance when we picture the temperature of a gas as determined by the kinetic energy of its molecules. Now I have no doubt that negligence of this sort can be justified, but the necessary logical analysis is apparently complicated, and involves a great many different sorts of experiment by methods of asymptotic approximation, by which we establish the existence of various sorts of physical constants, such as constants of emission and absorption and reflection and scattering and fluorescence and thermal conductivity. We do not need to make the analysis here, but I believe that some time it would be worth while to attempt it. Such an analysis will justify the principle so often used: that if a body is in thermal equilibrium the various processes involved, such as radiation or thermal conductivity, must when taken separately also be in equilibrium. Doubtless, if our experience had been confined to higher temperatures,

like that of the sun, this notion of different mechanisms acting independently would have been more difficult to acquire.

We next consider another fundamental concept of thermodynamics, that of quantity of heat. We are at first perhaps inclined to think of this as a comparatively straightforward concept, given immediately in terms of experience, but an analysis of the operations by which we measure quantity of heat will show that the situation is really most complicated. Consider, for example, Joule's experiment in which the mechanical equivalent of heat was measured by determining the rise of temperature of the water in a container when stirred by paddles driven by a falling weight. We do not question that the rise of temperature of the water has its origin in the mechanical work done on it by the paddles. But what about the rise of temperature of the container? We shall doubtless say that part of this rise comes from heat communicated to it by the warmer water in contact with it, and part from mechanical work done on it by turbulent impact of the water. But by what operations shall we measure what part of the energy communicated to the container is heat and what part mechanical work? We try to give an idealized answer to this question in terms of Maxwell demons stationed at all parts of the boundary of the containing vessel with small scale measuring instruments. To measure the heat entering at any point I can see nothing else for the Lilliputian observers to do but to determine the temperature gradient at

every point of the boundary from temperature observations at two different levels, and calculate the heat inflow from the gradient and the thermal conductivity of the material of the walls—there seems no way of measuring a flow of heat as such. The inflow of mechanical energy must be calculated from a detailed knowledge of the elastic waves and other large scale deformations of the walls. Here again there is an arbitrary element in our procedure; if our mechanical measuring instruments are on too gross a scale, we may miss mechanical energy which we would catch with finer instruments.

This situation which we have just submitted to detailed analysis is, I believe, typical of the general situation; it is not possible in the general case to find anything which we can call heat as such. Without further explicit examination, we can unambiguously speak of a body losing or gaining heat only when there has been no energy interchange of any other sort with other bodies. In such a case the heat is measured in terms of the temperature change of the body. The heat concept is in the general case a sort of wastebasket concept, defined negatively in terms of the energy left over when all other forms of energy have been allowed for.

The essential fact that a quantity of heat can by itself be defined only in terms of a drop of temperature is somewhat obscured by the usual method of thermodynamic analysis. In describing a Carnot engine, for example, it is specified that the engine shall work between a source and a sink so large that

their temperature is not changed by the heat given out or absorbed by them, so that the impression is natural that heat may in some way be measured apart from temperature changes. This of course is not the case; we merely require that the source and sink be so large that their temperature changes are of a different order of magnitude from those in the working substance itself, so that with respect to the working substance, source and sink may be considered to be at constant temperature.

Assuming now that we are able to measure quantity of heat in those cases in which the concept has meaning, let us examine the first law of thermodynamics, which we write in the form:

$$dQ + dW = dE$$

Here  $dQ$  is the heat imparted to a given body by other bodies,  $dW$  is the work of all kinds done on it from outside, and  $dE$  is the increase of internal energy. Now if this equation says what appears at a naïve first glance, it should say that we find experimentally that the relation written always holds between the measured quantities  $dQ$ ,  $dW$ , and  $dE$ . We have seen that in the general case it is not possible to assign a unique operational significance to  $dQ$  and  $dW$ , and presumably not to their sum. We ignore for the present difficulties of this kind and confine attention on  $dE$ ; how shall we measure it? I believe it does not take much examination to convince us that there are no physical operations for measuring  $dE$  as such, and that therefore the equa-



tion expressing the first law must have a different significance from that which appears on the surface. This is often recognized in the statement that the essence of the first law is that  $dE$  is an exact differential determined only by the variables which fix the internal condition of the body, and not a function of the path by which the body is carried from one condition to another. But what shall we mean by internal condition, and how shall we be sure that we have found all the variables required to specify it completely? Internal condition may be a most complicated thing and require many variables, as shown by a piece of iron with a complicated magnetic history or by a piece of aluminum about to undergo recrystallization after overstrain. Here again I believe there is no physical procedure by which general meaning can be given to this concept of internal condition. In specific cases we can state what the variables are which determine internal condition, and the criterion that we have found the correct internal variables is that  $dE$  shall be a complete differential in terms of them. The first law of thermodynamics properly understood is not at all a statement that energy is conserved, for the energy concept without conservation is meaningless. The essence of the first law is contained in the statement that the energy concept exists (or has meaning in terms of operations).

The first law is often thought to be one of the most general of physics, but in a paradoxical sense it is the most special of all laws, because no general

meaning can be given to the energy concept, but only specific meaning in special cases. The first law owes its complete generality to the fact that no specific case has yet been found of so broad a character that it cannot be included under one or another special case.

Examination will at once justify this view. Thus we find a great many systems which are adequately described in terms of two variables, pressure and temperature, in that a function of  $p$  and  $t$  can be found such that its differential equals  $dQ + dW$ . There are other systems in which the six components of stress and  $t$  completely fix the internal condition in the sense that they determine a  $dE$ . In other systems the specification of a magnetic field may be necessary, or an electric, or a gravitational field. No case is known which cannot be handled in terms of the action of external forces of the proper kind, but there is no general procedure, and the first law owes its generality to the exhaustive cataloging of special cases.

We may now return to the question left in abeyance above of the ambiguity in  $dQ + dW$ . In all the cases in which the specific variables can be found which define  $dE$ ,  $dQ$  and  $dW$  also have meaning. Consider, for example, a gas, the internal condition of which may be characterized in terms of  $t$  and  $p$ . The mere fact that the internal condition can be specified in terms of two variables, one a mechanical variable, shows that the substance is mechanically homogeneous. Being mechanically homogeneous, we do

not have the possibility of ambiguous values of  $dW$  varying with the scale of the measuring instruments, and in fact we know that  $dW = p \, dv$ . Similarly the gas being homogeneous and at rest as a whole allows unique values for  $dQ$ . Of course this cannot obscure the physical fact that even in such a gas, when we go to a small enough scale, we find inhomogeneities arising from the Brownian movement, etc. Practically our statement means that the inhomogeneities are so fine grained that over a very wide range of scale of the measuring instruments we find the same definite results. The same sort of considerations apply to more complicated systems. If  $dE$  is a complete differential in terms of  $t$  and six stress components, this means again that the body is homogeneous, its condition is determined by temperature and stress, which are the same throughout the body, and again there is no possible ambiguity from the scale of the instruments which measure  $dW$  and  $dQ$ . It seems in general, then, that if the body allows operations by which  $dE$  acquires meaning, at the same time  $dQ$  and  $dW$  are provided for. In working out this idea in full detail, some care must be given to the question of order of differentials.  $dQ$ , for unit time and unit volume, is strictly equal to  $k \nabla^2 t$ , where  $k$  is thermal conductivity, so that in determining  $dQ$  the second derivatives of temperature are involved.

If the body is obviously not homogeneous, it is still a matter of experience that it can be divided into small pieces, each of which are by themselves suffi-

ciently homogeneous, and the first law in its usual form may be applied to each of the pieces.

Finally, we emphasize a fact already implicitly mentioned, namely, that no physical significance can be directly given to flow of heat, and there are no operations for measuring it. All we can measure are temperature distributions and rates of rise of temperature. As at present defined, a heat current is a pure invention, without physical reality, for any determined heat flow may always be modified by the addition of a solenoidal vector, with change in no measurable quantity. If someone states that throughout all space there is a uniform heat current of  $10^6$  cal./cm.<sup>2</sup> sec. in the direction of Polaris, no disproof can be given, for such a stream is solenoidal, and as much heat flows out of every closed surface in unit time as flows in. Such a solenoidal flow is meaningless in terms of operations; we could give meaning to such a flow only in terms of some slight modification of the solenoidal condition introduced by the measuring instrument. In all ordinary conditions the flow of heat given by the simple relation  $q=k \text{ Grad } t$  corresponds exactly to what our atomic pictures lead to expect in those cases where the details of the picture can be worked out. But there may be cases where it is advantageous to supplement the ordinary heat flow ( $=k \text{ Grad } t$ ) by the pure fiction of a solenoidal flow, because in this way it may be possible to account for new phenomena which appear when the solenoidal conditions are slightly departed from. Thus if in a conductor at uniform temperature

carrying a steady electric current we say that a heat current is also flowing proportional to the electric current and therefore solenoidal, we may provide the possibility for a simple correlation of phenomena found under those more complicated conditions when an electric current flows in a conductor of non-uniform temperature in a magnetic field. If it should turn out that the heat current is *uniquely* determined by considerations of this character, then we have taken the first step away from the pure formalism which this sort of thing otherwise is in the direction of giving physical reality to the invention of "heat current."

There are other interesting questions of a fundamental thermodynamic character, such for example, as whether the entropy concept has any general significance apart from the scale of our measuring instruments, and what is the operational significance of applying thermodynamic concepts to radiation, but we shall not consider these questions here.

### ELECTRICAL CONCEPTS

We now set ourselves the problem of finding the meaning of the various concepts in terms of which we describe the behavior of electrical systems, assuming that we understand what we mean by "electrical." We start with the simplest electrical systems, namely, those in which we deal with static phenomena on a large scale. In such systems there are independent physical operations by which we may find the magnitude of any charge, provided that it is effectively



concentrated in a geometrical point. The measurements involved in these operations are measurements of ordinary mechanical forces; we assume that our knowledge of mechanics has already taught us how to make such measurements. An electrically charged body experiences forces, which may be measured by tying a string to it and pulling on the string with a spring balance hard enough to keep the body in equilibrium. Three charges are numerically equal if when each is placed at unit distance from another, in the absence of the third (or other charge), the forces are always the same. If furthermore the forces are of unit magnitude, the charges are defined as unit charge. Having obtained unit charge, we define the magnitude of any other charge as equal to the force which it experiences when placed at unit distance from unit charge. This of course is all very trite; the important thing for us is merely that magnitude of charge, or quantity of electricity, is an independent physical concept, and that unique operations exist for determining it. These operations presuppose the ability to perform certain operations of mechanics. Having now learned how to measure electrical quantities, we discover experimentally the inverse square law of force, and later arrive at the concept of the electric field. As we have seen, the field is an invention; here we shall use this concept only for the purpose for which the invention was made, and shall not involve ourselves in any of the implications of ascribing physical reality to the field. Notice that as long as we deal only with point charges we do not have to

define field strength in terms of the limiting procedure of making the exploring charge smaller, for the limiting small charge is necessary only to avoid the reaction of the exploring charge on the positions of the charges which generate the field. All this again is trite; the important point is that the operations by which the inverse square law and the concept of the field are established presuppose that the charge is given as an independent concept, since the operations involve a knowledge of charges. The operations also involve the measurement of forces by the ordinary *static* procedure of mechanics with spring balances. With the means now at our command we establish one very important property of electric charges, namely that the total amount of charge on an isolated body of finite size is conserved, no matter how the charge is forced to rearrange itself by the motion of charges on adjacent bodies.

By procedures exactly like those outlined above, we may treat all the corresponding magnetic quantities; there is formal parallelism between the two sets of phenomena, but there is the physical difference that we have to realize a single magnetic pole by the device of using a very long slender magnet.

We now give our electrical system more freedom, in that we allow the charges to be in motion with respect to each other. Perhaps the most immediate question which we now have to ask is whether charge continues to be conserved when set in motion, or whether the total charge on an isolated body is a function of its velocity? To answer this question we

must generalize the procedure by which we assigned a numerical value to a stationary charge. Perhaps the simplest way is to allow two unit charges each to move with constant velocity, remaining at unit distance apart, and measure with a spring balance the force required to keep them at constant distance apart. Now we immediately find that the force is altered under these conditions, so that our first impulse is to say that the charge is a function of the velocity. But as we experiment further, we find that the state of affairs is very complicated; the force between the two charges at any moment of their motion depends not only on the charges, their distance apart, and their velocities, but also on the angle between the line joining them and the direction of motion in the lines. Further experiment of other kinds yields other information; it requires a force to maintain a charge in uniform motion in a magnetic field, or to maintain a magnetic pole in motion in an electric field. A moving electric charge exerts a force on a stationary magnetic pole, so that by definition the moving charge is surrounded by a magnetic field, and similarly a moving magnetic pole is surrounded by an electric field. Returning to our two moving electric charges, we are impelled to ask whether, if all these complications are possible, the numerical constant (unity for static charges) in the inverse square law of force is a function of velocity as well as the magnitude of the charges themselves? If we broaden the question in this way, as we apparently must, our problem becomes indeterminate, for

we are trying to answer two different questions with a single kind of measurement, namely of the force between moving charges. I have had no better luck on trying other methods of measurement. Apparently the operations do not exist by which unique meaning can be given to the question of whether the magnitude of a charge is a function of its velocity. On realizing this situation, we are at first embarrassed to know how to proceed, but we reflect that the embarrassment is not of our own making, but corresponds to a physical fact. The concept of charge as a unique and independent thing essentially pertains only to static systems. We may extend the concept to moving systems if we wish, as a matter of convenience to ourselves, but must recognize that such an extension is an invention of ours and not a reality of nature. Now we do make such an extension, and we make it in the simplest possible way, that is, we define the charge on an isolated body in motion as that which we should find on it if we reduced it to rest and made measurements according to the regular static procedure. That this is a convenient thing to do depends on the experimental result that the charge so found is independent of the way in which velocity is imparted to or removed from the body; in other words, whenever the body is reduced to rest, the same charge is always found on it.

Although this is pure definition on our part, it turns out to have a most simple and convenient connection with experimental facts which were discovered after the decision to treat a moving charge in

this way was made; the discovery is of the atomic structure of electricity. If then we agree to call each elementary charge a constant independent of the velocity, the total charge on a body becomes merely proportional to the count of the total number of atomic charges on the body, which is certainly highly convenient and suggestive.

Having now fixed what we mean by the magnitude of a moving charge, we are ready to turn to the general problem of the behavior of any system of charged bodies in motion. For the present we consider only phenomena of the scale of everyday experience. The most general problem that has meaning here is to determine all measurable properties of the system in terms of those data which experiment shows can be arbitrarily specified. Now we have already emphasized that the electromagnetic field itself is an invention, and is never subject to direct observation. What we observe are material bodies, with or without charges (including eventually in this category electrons), their positions, motions, and the forces to which they are subject. The forces are to be measured according to definition in mechanical terms, either by the strains in members of a framework if the system is in equilibrium, or in terms of accelerations and masses if it is not in equilibrium. The electromagnetic field as such is *not* the final object of our calculations, but the calculation of it is only an intermediate auxiliary step, convenient to make because our mathematical formulation gives so simple a connection between electromagnetic field, charges,



and mechanical action that the latter can be calculated at once in terms of the former. In fact the connection is so simple that in many cases we have come to regard our problem as solved if we can compute the electromagnetic field, overlooking the fact that the field has no immediate meaning in terms of experience.

Electromagnetic theory now presents us with a solution of the general problem; this solution is contained in the four-field equations of Maxwell, the constitutive equations, and those additional equations (quite often lost sight of) which give the forces exerted by the field on electric charges, or currents, or dielectrics. Let us inquire how we may set about testing the physical correctness of these equations. We may begin with one of the simplest possible tests, and inquire whether the equations are correct in stating that the force acting on a charge moving in an electric field is simply the product of the charge and the field strength. This, on the face of it, is a surprising statement. The field itself is affected by the motion of the charges which generate it, and it is natural to expect a converse effect. If, furthermore, we have sympathy with the medium point of view, it is easy to think that whatever it is in the medium that gets hold of a charge and exerts a force on it will find it harder to take hold when the charge is in motion.

In attempting to check our statement experimentally, the only additional complication, as compared with the static case which we have already checked,

is afforded by the motion of the charge, for we have defined the magnitude of a charge in motion, so there is no difficulty here, and we may furthermore suppose that the field is generated by stationary charges, so that we need not trouble to inquire whether the procedure by which the field was originally defined is here applicable. The task of checking the equation then reduces to the simple physical task of measuring the force on the moving charge. How shall we do this? If the velocity is low, we may tie a string to the charge and measure the force with a spring balance (or its equivalent). But now an examination of the equations shows that in more complicated phenomena perceptible deviations from the static behavior are to be expected only at much higher velocities than can be attained by towing charges with a string and a spring balance, so that it is evidently necessary to check the simple equation for the force on a moving charge also at high velocity. Since at high velocity the spring balance method for measuring forces fails, we are driven to the only procedure that we have, namely a measurement in terms of the resultant acceleration, calculating the force by Newton's first law of mechanics. But this involves a knowledge of the mass of the moving body, which we recognize in general may be a function of the velocity. Now we have already seen, in discussing the concepts of mechanics, that the operations by which mechanical mass is defined cannot be carried out at high velocities, so that either the concept of mechanical mass becomes meaningless at high

velocities, or we must adopt another definition. In attempting to give this new definition of mass at high velocities, we are driven to a result of special relativity theory, namely that all mass, mechanical or electrical, must be the same function of velocity. If now electrical mass can be found in terms of velocity, our immediate problem is solved and we shall be in a position to complete the experimental check of the equation. But as a matter of fact, in order to determine electrical mass, we have to use that equation which we are now engaged in trying to establish. Logically we have again the vicious circle, the physical significance of which is that independent operations do not exist for giving unique meaning to the concept of force on a charge at high velocity.

We seemed so close to our goal a minute ago; that we may allow ourselves to jump the logical chasm, and assume that the equation is correct. Electrical mass now becomes a definite function of velocity, mechanical mass the same function, and we are in a position to compare the actual acceleration received by a charge in a field with that calculated by the equation. Our conviction, on the basis of all experience up to the present, is that the two accelerations will be found to agree.

The equation then does somehow make correct connection with experience in that a consequence of the equation can be verified experimentally, in spite of the fact that as the equation stands it is meaningless, because the operations do not exist by which meaning can be given to the individual terms. At low veloci-

ties the equation really says what it seems to say, because the individual terms have meaning in terms of operations; and, what is more, what the equation says agrees with experiment. At high velocities the equation does not mean at all what appears on the surface; by itself it has no meaning; it has meaning only when considered as a member of a system of equations, and only in so far as the system of equations makes by implication statements about nature that have meaning in terms of operations that can be carried out physically. The individual terms of the equation of the system do not have meaning at high velocities, and in fact there are more terms than there are independent physical operations.

An exact analysis from the operational point of view of the significance of the equations at high velocities has perhaps never been made, and is not necessary for our immediate purpose. The discussion has brought out, however, that the number of physically independent concepts has been cut down by two at least, in that we have made purely formal definitions of the meaning of quantity of electricity, and of the force exerted by a field on a charge at high velocity. There is no reason to think that there is anything unique about this analysis, or that formal definitions might not have been given to other concepts than charge and force. We can only state that as far as physical content goes the equations have at least two degrees of freedom. It should then be possible to find quite different sorts of equations which agree equally well with experience. In particular,

since we have seen that the force on a moving charge has no meaning in terms of independent operations, it should be possible by arbitrary definition to make this force any function of velocity that we please (of course reducing to the proper value at low velocities), and then to determine the other equations so that the entire group of equations is consistent with experiment. So far as I know, no one has tried to give such a modified set of equations, and indeed there is no particular reason why anyone should bother to do this, because the present equations are simple enough, and the modified equations, although perhaps differing greatly in appearance from the present ones, would have no advantage in any greater or different physical content.

But there is no reason to think that the present state of affairs will always continue. We have seen that the decrease in the number of concepts corresponds to our inability to measure as many sorts of physical things at high velocities as at low. Now it is the task of the future experimenter so to refine the possibilities of measurement at high velocities as to restore these two degrees of freedom. In particular, mass should be made measurable in mechanical terms at high velocities. When this restoration has been made, and all the quantities in our equations receive independent physical meaning, the significance of the equations in terms of operations will be quite altered, although the formal appearance will be unchanged. We must then be prepared to find, as always when we change the range of phenomena, that the equa-



tions in their present form do not correspond to the facts at all, and that one of the alternative forms allowed by our present two degrees of freedom is the correct form. But until the new experimental facts have been obtained, it seems hardly worth while to attempt to specify the doubly infinite variety of forms which the equations might have consistently with present experiment.<sup>1</sup>

So far we have discussed the extension of ordinary electric phenomena in only one direction, to high velocities. There is another extension which is much more important physically, namely to very small scales of magnitude. This extension is necessary to an understanding of the properties of matter in bulk, the electrical nature of the atom having been once established. Our problem is to show how the statistical average of the behavior of a large number of electrons gives the large scale effects which are within the reach of observation, and which are described by the equations we have just discussed. To get this statistical average we must be able to calculate at least certain features of the behavior of the individual electrons, which means that we must know the form of the equations down to dimensions of the order of those of an electron, or smaller. Now if one contrasts the scale of the supposed dimensions of the electron with the smallest dimensions on which we can make

<sup>1</sup> Since this was written, a paper has appeared by V. Bush, *Jour. of Math. and Phys.*, vol. V., No. 3, 1926, in which it is shown that there are advantages in supposing the charge of an electron to change when it is set in motion.

independent experimental verification of those equations, he must admit that there is an enormous chance for change in the type of equation beyond the limit that we can reach by direct experiment, and the chances of guessing the correct extension of the equation to small dimensions are correspondingly almost vanishingly small. (We may perhaps say that experiments on the Brownian motion on a scale a good many atoms in diameter bring us the closest possible directly, which means that we are  $10^8$  or  $10^7$  fold away from electronic dimensions.) In spite, however, of the apparently enormous chances against it, this program of extending the field equations to small dimensions and following out the consequences was exactly the program which Lorentz set himself.<sup>1</sup> That Lorentz saw that such a program might be carried through must be recognized as a vision of extraordinary genius, and that he was willing to devote to it the years of arduous and detailed calculation that he did is evidence of a pertinacity of purpose of the highest moral order.

We now have to examine critically this program and to inquire what is the significance of the measure of success that Lorentz attained. The precise extension of the equations that he made was very simple, for the large scale equations of Maxwell were taken over with as little change as possible. The equations are so familiar that it is not necessary for us to write them in detail; they express relations between the

<sup>1</sup> See for example, H. A. Lorentz, *The Theory of Electrons*, B. G. Teubner, 1916.

electric and magnetic force vectors (force and induction now becoming the same thing, the difference between them in ponderable bodies being one of the things that is to be explained in terms of the electrons), the space density of electric charge, its velocity, and the force acting on elementary charge. We have to notice that although formally the equations have changed little in appearance, nevertheless the physical content, as judged by the operations, has changed a great deal. Consider, for instance, the meaning of charge density. In the Maxwell equations,  $\rho$  was merely the number of discrete elementary charges per unit volume, the distances between these charges being supposed so small compared with the scale of the phenomena involved that their average effect could be fairly represented in terms of their numbers. In the Lorentz equations, on the other hand,  $\rho$  has a value different from 0 only inside the electron; everywhere else  $\rho=0$ . Now an examination of the previous discussion, in which we questioned whether the magnitude of the charge might be a function of its velocity, will show that there are no physical operations whatever by which meaning can be given to  $\rho$  at individual points inside an electron. There is a single condition on this  $\rho$ , namely, that its integral throughout the total volume assigned to the electron shall equal the total static charge of the electron. Obviously a single scalar condition is a pretty blunt tool with which to attempt to determine a point function throughout a volume. Again, the

equations talk about the velocity of the charge at interior points of the electron; what possible physical operations are there by which meaning can be assigned to the velocity of an amorphous structureless substance in regions inaccessible to experiment? Here again, the concept as a detailed description of the behavior at a point has become meaningless, and again there is a single integral condition, namely, that the  $v$  associated with every  $\rho$  must be such that when integrated over the volume of the electron it will give a total transport of charge equal to that carried by the electron in its motion. This again is a single condition on a function distributed through space. Still again, the equations contain the electric and magnetic vectors at points inside the electron. What is the possible meaning of these field vectors in terms of operations? Our procedure for finding the field at a point involves by definition finding the force on an electric charge placed at that point. But there is no charge smaller than an electron, and the procedure degenerates into a fiction. Again there is a single integral condition on the field vectors; the integral of the force on the assumed charge density when taken over the total volume of the electron must give a value corresponding to experiment. Except for this single condition, the concept of the field at points inside the electron is an invention without physical reality. Not only is the field concept meaningless at points inside the electron, but it is meaningless at points outside within a certain distance, because the exploring

charge can never be made smaller than the electron itself, and so can never come closer than a certain distance.

The actual state of affairs is much worse than has already appeared. It was shown in the discussion of space and time that no independent physical meaning can be attached to lengths and times as small as have to be assumed in describing the behavior of the individual electrons. The operations Div, Curl,  $d/dt$  which enter the field equations are, therefore, physically meaningless as they stand; they have only a mathematical meaning which begins to acquire physical complexion in a most complicated way when the equations are integrated over large enough volumes.

It is evident, therefore, that the concepts which enter the field equations have entirely lost their large scale significance; they have become blurred, fused together, and fewer in number. A precise analysis of this situation has probably never been attempted and would obviously be difficult: it would be interesting to know at least how many really independent concepts there are at this order of phenomena. An attempt at an analysis would probably be worth while from a physical point of view in suggesting possible experiments by which the number of physically independent concepts could be extended.

Since the quantities in the field equations are meaningless in the naked form in which they enter the equations, it is meaningless to inquire whether the equations as they stand are true or not. In our pres-



ent state of experimental knowledge it is also meaningless to ask whether, for example, the inverse square law between electric charges continues to hold, or whether an accelerated charge radiates. These questions have meaning only when applied to phenomena on a scale large enough to correspond to possible experiment.

There is a rather interesting obverse to the statement that it is meaningless to ask whether the field equations are true, namely, that it may not be meaningless to state that they are false. A statement is not true unless it is true in every particular, but it is false if it is false in a single particular. If we can show that a single consequence of the field equations of Lorentz, when integrated or averaged in such a way as to correspond to experimental possibilities, is false, then the equations must be false. It seems that, regarded as a complete description of physical behavior on a small scale, the equations must be judged false, because they contain no suggestion of quantum phenomena.

Even if we have to recognize that the equations are false, there can be no question that they correspond to an important part of reality, and that they have been of the greatest service to physics. What is the significance of the success that they have attained? It is to be noticed that all the phenomena to which the Lorentz equations have been successfully applied, although not large scale phenomena in the ordinary sense of the word, are nevertheless phenomena involving the coöperation of a number of atoms, and that

the equations unquestionably fail when applied to phenomena involving single electrons. It appears from our best present evidence that on a small scale the behavior of nature is governed by quantum principles and is therefore quite different from large scale behavior, which we have seen is governed by the Maxwell equations. There must of course be a transition zone in which the character of phenomena changes from quantum to Maxwell. Now any program like that of Lorentz is almost inevitably bound to begin to give correct results when we get up as far as the transition zone, for the simple reason that the relations of Maxwell have been put into the equations and are always there ready to appear as soon as the quantum relations begin to give way. The physical significance of the success of the Lorentz program seems to be that the transition from Maxwell to quantum takes place at a stage pretty far down toward the individual atoms. To find the precise details of the transition from Maxwell to quantum phenomena constitutes a large part of the program of the immediate future.

All this skepticism about the classical work of Lorentz is likely to be rather irritating or depressing, particularly if one attempts to imagine what other course could have been adopted. Indeed it does seem that we find ourselves in a real quandary; Lorentz was practically forced, because of the character of the mathematical tools at his command, to take the course that he did, in spite of any recognition of the physical meaninglessness of the mathematical operations. We

have already seen that conventional mathematics does not correspond to the physical reality; it cannot easily make a qualified statement subject to limitations, and it recognizes no difference between the physically big and the physically little and the corresponding change in the operational meaning of its symbols. It begins by being a most useful servant when dealing with phenomena of the ordinary scale of magnitude, but ends by dragging us by the scruff of the neck willy nilly into the inside of the electron where it forces us to repeat meaningless gibberish. Larmor recognized this, and in his electron theory, developed practically contemporaneously with that of Lorentz, endeavored to treat electrons as wholes, and not to make meaningless statements about their insides.<sup>1</sup> But he was much less successful than Lorentz in making his analysis give physical results, and one may suspect that it was at least in part due to difficulty with his tools.

What we should like to be able to do is easy to see. The things that go into our equations must have independent physical meaning, and the character of our mathematical formulation should change to keep pace with the change in the physical operations which give meaning to the terms. For example, electrical density has a meaning for large scale phenomena, but means nothing on a small scale. Our ultimate electric unit is the electron; when we get down to this scale of magnitude, our mathematics ought to be making

<sup>1</sup> Joseph Larmor, *Æther and Matter*, Cambridge University Press, 1900. In this book the electron is treated as a point singularity in the ether.

statements about the relative behavior of discrete electrons, and not mention so much as by implication the density at points inside an electron. But this sort of thing we apparently cannot yet do; the proper mathematical language has not been developed. Such a language, when developed, must not only be able to resist the temptation to burrow inside the electron, but must also try to get along without the field concept, which we have seen is liable to so much physical abuse, and must reduce effects in complicated electrical systems to the ultimate elements that have physical meaning, namely, a dual action between pairs of electrical charges, with no implications about physical action where the charges are not.

#### THE NATURE OF LIGHT AND THE CONCEPTS OF RELATIVITY

We have already discussed several aspects of the theory of relativity in connection with the relation to it of some of our fundamental concepts. There are still other topics connected with relativity which demand attention; most of these involve the properties of light. It will now be convenient to discuss together the properties of light and these concepts of relativity. We restrict our discussion of light to those simple properties which bear on the theory of relativity.

Practically all our thinking about optical phenomena is done in terms of an invention, by means of which these phenomena are assimilated to those of ordinary mechanical experience, and so made easier

to think about. To realize that invention has been active here, we must think ourselves back into that naïve frame of mind in which experience is given directly in terms of sensation. The most elementary examination of what light means in terms of direct experience shows that we never experience light itself, but our experience deals only with things lighted. This fundamental fact is never modified by the most complicated or refined physical experiments that have ever been devised; from the point of view of operations, light means nothing more than *things lighted*. Now experience shows that these things lighted may stand to each other in varied relations; in attempting to reduce these relations to order and understandability we make a certain invention. This is prompted by several cardinal experimental facts: in the first place, things lighted have a simple geometrical relation to each other, in that screens placed on straight lines between the lighted objects may suppress the illumination of one or the other and themselves become illuminated. This leads to the concept of rectilinear beams of light, which is no more than a description of the geometrical relation between lighted objects. Then we have the experimental fact of the asymmetrical relation of the lighted objects, described in terms of sources and sinks. Finally, we have the discovery made at a much later stage, and not possible until physical measurements had reached a high refinement, that light has properties analogous to the velocity of material things. This was first discovered in connection with astronomical phenomena



in the shift of the time of eclipse of Jupiter's satellites and in aberration, but was later found to hold for purely terrestrial phenomena, in that a beam of light reflected from a distant mirror does not return to the source until after the lapse of a time interval that can be measured with means sufficiently refined. This property of return after the lapse of time is precisely like that of material things, such as a messenger despatched for an answer, or a ball or a water wave bouncing from a wall. These various properties of light lead quite naturally and almost inevitably to the invention of light as a thing that *travels*, "thing" not necessarily connoting material thing.

The question now for us is whether we shall regard this as a mere invention, made for convenience in thinking, or shall go further and ascribe physical reality to it, that is, shall we think of light as capable of independent physical existence in the space between the matter that constitutes the source and the mirror? Now in spite of the resemblances pointed out above, there is at least one universal and fundamental difference between a thing that travels and light. We have independent physical evidence of the continued existence of the ball, for example, at all intermediate points of space; we can see it, or hear it, or feel the wind in the air as it passes, or even touch it. All these phenomena are independent of the initial and terminal phenomena, and hence by our criterion for the physical reality of an invention, we are justified in ascribing physical reality to the ball in transit. But with the beam of light it is entirely different; the

only way by which we can obtain physical evidence of the intermediate existence of the beam is by interposing some sort of a screen, and this act destroys just that part of the beam whose existence we have thereby detected. There is no physical phenomenon whatever by which light may be detected apart from the phenomena of the source and the sink (understanding that a mirror is included in the idea of sink); that is, no phenomenon exists independent of the phenomenon which led us to the invention of a thing travelling. Hence from the point of view of operations it is meaningless or trivial to ascribe physical reality to light in intermediate space, and light as a thing travelling must be recognized to be a pure invention.

The status of light is exactly the same as that of an electric field; there is not the slightest warrant for ascribing physical reality to either at points of empty space—light and field-at-a-point have no meaning until we go there and make experiments with some material thing. Of course the electromagnetic theory of light makes this resemblance inevitable, provided the theory and our views of the nature of light and the field are correct.

It cannot be denied that there are some phenomena which uncritically considered appear to justify thinking of light as a thing that travels; these will now be discussed. Probably the argument to which most significance is usually ascribed is derived from the phenomena of energy. The passing of light from source to sink is accompanied by the transfer of

energy. But energy is conserved, so that we have to ask where the energy is in the time interval between the emission of light from the source and its absorption by the sink. There is an obvious answer: the energy is in transit, of course, somewhere in the intermediate space between source and sink. If we think of light as propagated through a medium, then the medium is such that energy may reside in it, as in the electromagnetic theory of light, or if light is more material and ballistic in character, the thing that travels has itself energy. We notice in the first place that the conservation principle involves the time concept, because what we mean by conservation is that the total energy of the universe, at a fixed instant of time, is constant. That is, we have to integrate over all space the local energy at a definite instant of time, and this involves spreading the time concept over all space. It is further evident that unless we spread the time concept over space in the right way we shall not get conservation. The proof that it is possible to spread the time concept over space in such a way as to give conservation involves a knowledge of the properties of light. It would seem, then, that we ought not to assume conservation in deducing the properties of light, when a knowledge of the properties of light is necessary to establish conservation. These considerations cannot be accepted as final, however, until a detailed analysis has been made, and this would be most complicated. But there is a more important consideration derived from our previous critique of the energy concept, namely, that there is no basis for

asserting that energy is localized in space at all; energy is not a physical thing, but rather what we would call a property of a system as a whole. If this view of energy be granted, the whole energy argument for light as a thing travelling, and also for the existence of a medium, falls. I believe that similar considerations apply to any arguments from the conservation of momentum.

The possibility of detecting light in apparently empty space by a screen constitutes perhaps the most immediate reason for considering light as a thing that travels. This point of view I believe is characteristic of the entire attitude of Einstein in deducing the theorems of the special theory of relativity. Einstein's light signal is for the purposes of the deduction thought of as a simple spherical wave spreading from the source and capable of being watched as it spreads by an observer outside the system, in much the same way that a water wave can be watched. Of course the light signal cannot actually be watched in transit, but we can come fairly close to this ideal by placing screens at any point we please to make the wave visible. It is true that the mere act of showing the existence of the light destroys that part of the beam whose existence is detected, but the screen needs only an infinitesimal amount of light to make it visible, and so by the usual physical argument we may suppose that the detecting screen produces only an infinitesimal modification of the total original light.

Our satisfaction with this picture evaporates if our present quantum views of the nature of light are cor-

rect. We can no longer think of the spherical light pulse as of irreducible simplicity, but it is an exceedingly complicated thing, perhaps more complicated than a gas from the point of view of kinetic theory, and simulates simplicity by some sort of averaging of the effects of the elementary quantum processes of which it is composed. If the principles of relativity are to continue to be regarded as fundamental, or even if they are to remain intelligible, we must apply our reasoning, not to spherical wavelets, but to the elementary process of which these wavelets are composed. Now the elementary quantum act is essentially a twofold thing: there is a discrete act of emission at some discrete material particle, and the act is consummated by another discrete act (absorption or scattering) at some other discrete particle. We cannot yet fully characterize the details of this twofold process, but have to connect the place at which absorption takes place with the place of emission by statistical considerations. It is evident, however, that to think of emission as starting some process like a spherical wavelet travelling like a thing through space presents an entirely incorrect view, because in the wave there is no hint of the discrete place which is to terminate it. We may say crudely that there is no way by which the wave can *know* what discrete material particle is to complete the emission process. We may perhaps try to save the situation by remembering that a spherical wave is polarized and so has a unique direction associated with it; but further examination shows that this does not help, because the



unique direction is that of no energy flow, and absorption can take place in any direction *except* this. It appears then that instead of being a help, the thing-travelling point of view is a positive embarrassment when we try to picture by means of it the essentially twofold nature of the elementary quantum act.

Another plausible argument for light as a thing travelling may be deduced from our principle of connectivity. Imagine a dark lantern with a shutter that can be opened or closed so as to emit a momentary flash of light, a distant mirror, and a receiving instrument near the source. One of the properties of light that we always assume is that no permanent trace of the act of emission is left behind in the source. The most minute examination of all the details of the lantern and its surroundings at some time after the emission of the flash has not yet shown any phenomenon that betrays a remembrance of the emission of the flash, unless perhaps we measure the total energy or momentum and have some way of knowing what the energy or momentum would have been if the flash had not been emitted, and in any event we cannot specify the moment in past time when the signal was emitted. In the same way we cannot tell from an examination of a mirror whether it has at any time in the past reflected a beam of light. Consider now two systems, each consisting of a source and a mirror distant  $3 \times 10^{10}$  cm., identical in all respects, except that in one a light signal was flashed from the source 1.5 seconds ago, and in the other only 0.5 seconds ago. According to our hypothesis, the most complete

examination of source and mirror in either system fails to show the slightest difference, but nevertheless there is something essentially different about the two systems, for in one a light signal arrives at the screen in 0.5 seconds, and in the other not until 1.5 seconds. This violates what we have suggested might be regarded as the cardinal and most general principle of all physics, the principle of essential connectivity, which states that differences between two systems must be associated with other differences. A most obvious and simple way of maintaining our principle is merely to point out that the system really included more than we investigated: the system properly consists of source, mirror, screen, and all intermediate space, so that if we had examined intermediate space we should have found light there in transit in different positions in the two systems, thus correlating with the differences in subsequent history. This argument appeals to me as perhaps the strongest that can be advanced for the view of light as a thing travelling. But it seems in no way conclusive. Our principle of essential connectivity made no mention of the time concept, but we have somehow smuggled it in in making the application above. We sought to give a complete description of our system at some one instant of time, and this involved spreading the time concept over space. This itself is a questionable operation and may be done in different ways. But, more important, what is the justification for supposing that the system can be completely described by giving a complete description of all the measurable parts of it at

some one epoch? We have seen that in the most general case the principle of essential connectivity must recognize that the concept of "initial condition" of a system involves all the past history, and we may have here a case in point. The answer can be given only by experiment. In dealing with ordinary experience, when we do not have to distinguish between local and extended time, and are not dealing with optical phenomena, there can be little question that experience at least approximately justifies the expectation that future behavior is determined in terms of the present condition and that present condition may be specified in terms of the results of present operations performed in the system. But before extending this principle to phenomena in which we have to distinguish between local and extended time, we have to answer just that question which we are now considering; namely, whether there are physical phenomena taking place in apparently empty space, and whether therefore empty space has to be included in the system. We find ourselves again treading the vicious circle. Perhaps experience will show that the extension of the principle of connectivity to optical phenomena involves something like this: namely, the future at any point in a material system is determined by a complete description of the *present* state of the system in the immediate vicinity, and by a *history* of the behavior at more distant points, this history extending over longer and longer intervals of time as the point becomes more remote.

I believe, however, that these possibilities will not

seem very satisfactory, and that most physicists will discover in themselves a very strong disposition to feel that the future is determined in terms of a complete description of some sort of instantaneous configuration, time being extended in some suitable way over space. This instinctive demand that the future be determined in terms of the present may easily be consistent with the optical phenomena in our two systems consisting of source, mirror, and screen, without involving the material existence of light in empty space, provided that our assumption that the emission of light leaves behind it no permanent record in the source was incorrect. It may be that detailed examination of a source after emission will disclose permanent traces, from which the instant of emission may be found by extrapolation. If the conviction of determinism of the future is strong, the physicist may well be impelled to search here for new phenomena indicating such a memory of emission.

Let us now inquire how our physical structure might be affected if we should give up the identification of light with a thing travelling. One consequence is that light need no longer be thought of as having the property of velocity, since velocity, in terms of immediate experience, is a property of things moving from place to place. Giving up the concept of light as a thing travelling would enable us, then, to adopt an alternative method of describing nature with a different concept of velocity; we have seen that it is possible to define velocity in terms of operations different from the usual ones, in such a

way as to give the usual numerical results at small velocities, but different results at high velocities, and in particular to give an infinite velocity for light.<sup>1</sup>

There is now no objection to an infinite number associated with light, if we no longer think of light as having physical velocity. We may, if we like, continue for the sake of convenience to talk of the velocity of light, clearly understanding that the infinite value which must be ascribed to this velocity corresponds to the fact that the physical concept of velocity does not apply in all respects to light. We should now have to revise our process for extending the time concept over space, because this was formerly so done as to give light a finite velocity. We are now to make this velocity infinite, which is obviously to be done simply by setting a distant clock on zero at the instant it receives a light signal flashed from our clock at its zero. The behavior of material things now takes on a simple aspect—there is no longer a finite upper limit to the velocity that can be given a material thing, and light has no longer the paradoxical property of bearing the same finite relation to each of two material systems which differ from each other by a finite amount (that is, the first postulate of relativity that the velocity of light is  $3 \times 10^{10}$  in all reference systems). Light instead now bears the relation of

<sup>1</sup> No difficulty arises from the asymmetric character of light in assigning an infinite velocity to light because those physical operations by which we discover which is the source and which the sink are entirely distinct from the operations by which a velocity is measured; or in other words, even in the limit, it still has meaning to say that an infinite velocity has a direction associated with it.



infinitude to each of two systems which differ from each other by only a finite amount, and this is natural from the mathematical point of view.

However, all is not simplified by this change in the method of setting clocks, but a price has to be paid. The price is that we have to give up the simple connection between the velocity of a thing and its "go and come" time. Our changes have not affected local time; the time of passage of light to a distant mirror and return to the source is not changed, and is therefore still finite, although we describe the velocity of light as infinite. Now examination shows at once that there is no immediate connection between the concepts of "velocity" and of "go and come" time, because the operations involved are different. A measurement of a linear velocity according to our definition involves two clocks at two different places or else a clock travelling with the object, while "go and come" time demands only a single clock at a single place, and also involves necessarily a reversal of direction of motion in the object under measurement. We see then that, according to the definition adopted for velocity, we have the choice either of doing as Einstein did in the restricted theory of relativity and making "go and come" time very simply related to velocity; or we may say that refined physical measurements show that something of significance happens when the direction of motion is reversed, and that phenomena are not symmetrical with respect to a reversal of direction. The asymmetry which results from reversing the direction of motion we

may visualize as a sort of curvature in space and time, as of a small piece of an arc of a circle bent back on itself, with the two ends diverging. This alternative way of treating velocity would mean that velocity can be measured simply only by a specially situated observer; this need not be considered disturbing, because in fact the operations have been defined only with respect to such an observer.

Which of these two possible treatments of velocity shall be adopted is to a certain extent a matter of convenience, determined by the sort of phenomena in which we are most interested and wish most to simplify. Einstein's chief concern was with optical phenomena, so that the motive for his choice is evident. In this choice of Einstein it is not very evident that the desire to make "go and come" time simply connected with velocity played a very prominent part, but it seems rather that the desire to think of light as a thing travelling, with a finite velocity, was much more influential. This way of thinking of light is fundamental to all the treatment of restricted relativity; without this sort of picture all the mathematical deductions would lose their simplicity and convincingness, for in all the deductions we inevitably think of ourselves as an observer from outside, watching a thing that we call light travelling back and forth like any physical thing.

Now there can be no doubt that, when choice is possible, convenience and simplicity are important considerations; but I believe that there is another much more important consideration, namely, the

most perfect reproduction possible of the physical situation. It seems to me that it is very questionable whether Einstein, and all the rest of modern physics, for that matter, have not paid too high a price for simplicity and mathematical tractability in choosing to treat light as a thing that travels. Physically it is the essence of light that it is *not* a thing that travels, and in choosing to treat it as a thing that does, I do not see how we can expect to avoid the most serious difficulties. Of course the whole problem of the nature of light is now giving the most acute difficulty. The thing-travelling point of view, even as treated by Einstein, does not land us in a situation which is at all satisfying logically. We are familiar with only two kinds of thing travelling, a disturbance in a medium, and a ballistic thing like a projectile. But light is not like a disturbance in a medium, for otherwise we should find a different velocity when we move with respect to the medium, and no such phenomenon exists; neither is light like a projectile, because the velocity of light with respect to the observer is independent of the velocity of the source. On the other hand, in aberration we have a phenomenon similar to that shown by projectiles. The properties of light are more like those of a projectile than is perhaps commonly realized, as is shown in the papers of La Rosa<sup>1</sup> on the ballistic theory of light. The properties of light remain incongruous and inconsistent when we try to think of them in terms of material things. Einstein's re-

<sup>1</sup> M. La Rosa, *Scientia*, July-August, 1924.

stricted relativity has made a great contribution in so grouping and coördinating the phenomena that they can all be embraced in a simple mathematical formula, but he does not seem to have presented them in such a light that they are simple or easy to grasp physically. The explanatory aspect is completely absent from Einstein's work.

In view of all our present difficulties it would seem that we ought at least to try to start over again from the beginning and devise concepts for the treatment of all optical phenomena which come closer to physical reality. No one realizes more vividly than I that this is a most difficult thing to do. If we are ever successful in carrying through such a modified treatment, it is evident that not only will the structure of most of our physics be altered, but in particular the formal approach to those phenomena now treated by relativity theory must be changed, and therefore the appearance of the entire theory altered. I believe that it is a very serious question whether we shall not ultimately see such a change, and whether Einstein's whole formal structure is not a more or less temporary affair.

Although it is exceedingly difficult to foresee what the treatment of the future will be like, it is easy to surmise certain of its features. In essence the elementary process of all radiation perceived as radiation is twofold. There is some process at the source and some accompanying process at the sink, and nothing else, as far as we have any physical evidence; furthermore, the elementary act is unsymmetrical, in

that the source and the sink are physically differentiated from each other. This is the most complete expression of the physical facts; there is nowhere any physical evidence for the inclusion of a third element (the ether). Therefore all the phenomena apprehended by an observer (and this embraces all physical phenomena) can be determined only by the source and the sink and the relation to each other of source and sink, for there is nothing else that has physical meaning in terms of operations. This formula covers not only the possibility of such first order phenomena as aberration and the Döppler effect, but also shows that such second order effects as that looked for by Michelson and Morley must be non-existent. It will thus be seen that some of the consequences of relativity theory are implicitly contained in certain very broad points of view. One interesting question that must be answered before we can get very far with a new treatment is whether the elementary optical process is of *necessity* twofold, or whether we may have emission without absorption, that is, radiation into empty space. Lewis seems to imply in recent papers that this is not possible.<sup>1</sup> The astronomers have already pointed out difficulties in explaining phenomena like the temperature equilibrium of the planets if we suppose this is the case.

### OTHER RELATIVITY CONCEPTS

We now turn to some of the other concepts of rel-

<sup>1</sup> For example, in the book: G. N. Lewis, *The Anatomy of Science*, Yale University Press, 1926, p. 129.



ativity. One of the most important of these is the "event"; in fact this concept is made fundamental by Whitehead.<sup>1</sup> We have already discussed the concept of "event" under the "identity" concept with which it is closely involved. The event is usually thought of by Einstein as merely an aggregate of four coördinates, three of space and one of time. The principle of general relativity, namely, that the laws of nature shall be of invariant form, when formulated mathematically, involves the assumption that nature may be analyzed into events, and is expressed by the requirement that the mathematical relations between the coördinates of a chain of events shall be invariant. The same idea is also expressed by Einstein in another form, namely, that nature may be completely characterized in terms of space-time coincidences. In elaborating this idea, Einstein assumes that the results of all measurement may be given in terms of such coincidences.

Now it appears to me most questionable whether the analysis of nature into events is possible or sufficient. With regard to the coincidence point of view, it seems perfectly obvious that the world of our immediate *sensation* cannot be described in terms of coincidences; how, for example, shall we describe in terms of space-time coincidence the photometric comparison of the intensity of two sources of illumination, or the comparison of the pitch of two sounds, or the location of a sound by the binaural effect? To

<sup>1</sup> A. N. Whitehead, *An Enquiry Concerning the Principles of Natural Knowledge*, Cambridge University Press, 1919, Chap. V.

justify the coincidence point of view we apparently have to analyze down to the colorless elements beyond our sense perception. It does not seem unreasonable, perhaps, to expect that the universe is completely determined in terms of the positions as a function of time of all the positive and negative electrons; but to introduce such a thesis now certainly goes beyond present experimental warrant, and is contrary to the general spirit of relativity, which nowhere else involves any reference to the small scale structure of things. Even if we were willing to overlook all these objections, we would still have the fact that the difference between a positive and negative electron is not contained in any specification of the mere coördinates.

A further very important doubt in principle as to the possibility of the analysis of nature into events is afforded by the character of the concept of event itself. We have seen that the idea of event involves the existence of discontinuities, and that this can correspond only approximately to the physical fact, because discontinuities apparently lose their abruptness as we make our measurements more refined. The thesis that nature can be described in terms of discontinuities of a very small scale seems much too special to be made a fundamental part of a theory of the general pretensions of that of relativity. In fact this, as well as a consideration to be mentioned later, suggests that the argument and result of general relativity may be intrinsically restricted to large scale phenomena.

We now pass from these somewhat special questions to ask why it is that Einstein was able in the general theory of relativity to obtain new and physically correct results from general reasoning of an apparently purely mathematical character. We are convinced that purely mathematical reasoning never can yield physical results—that if anything physical comes out of mathematics it must have been put in in another form. Our problem is to find where the physics got into the general theory.

There are two questions to be disentangled here: we have to consider in the first place the significance of the fact that Einstein has been able to describe relations in nature in mathematical form, and in the second place of the fact that he was able to arrive at the mathematical formulation of these physical relations by reasoning of apparently a purely mathematical character, from postulates of merely formal mathematical content (invariance of natural laws in generalized coördinates). Now the theory of relativity does not seem to differ in the first respect from any other branch of mathematical physics, such as the classical mathematical theory of electricity and magnetism, for instance, and this matter has already been touched in an earlier chapter. It is a fact that the behavior of nature can in many cases be expressed to a high degree of precision in mathematical language, and relativity is not unique in this respect. In any event, we must not allow this possibility of mathematical formulation to obscure the essential fact that all physical knowledge is by its nature only approxi-

mate, so that we may expect at any time to find, when we have carried our measurements to a higher degree of precision, that our mathematical expression of the laws was not quite exact, as seems now to be the case with Newton's law of gravitation, for example. I do not suppose that Einstein would claim that the statements of relativity differ in this respect from any of our other statements about nature, although apparently some of his followers see something more here. (From the operational viewpoint the meaning to be attached to "something more" is somewhat obscure.)

With respect to the second question, we may stop to notice that the special theory stands in quite a different position from the general theory. The special theory is much more physical throughout; its postulates are physical in character, and it is obvious that the physics got into the results through the postulates. It seems to me without question that Einstein showed the intuitive insight of a great genius in recognizing that there are mutual relations between physical phenomena which can be described in very much simplified language in terms of concepts slightly modified from those already in common use. In view of the remarks made on the nature of light, it is legitimate to wonder, however, whether the formulations of even the special theory will always stand. It seems to be true that *all* the facts of nature, even in the absence of a gravitational field, cannot be connected by the simple formulations of the special theory; that the physical relations are simple only in a sub-group; and that if we wish to deal with *all* opti-

cal phenomena, we have carried our simplifications too far, for the emission of a light signal is not a simple event, and light is not in nature like a thing travelling. Just the sorts of physical thing which are ignored in treating light as the special theory does are coming to be more and more important in the minds of physicists, and this is a reason for wondering whether ultimately Einstein's special theory may not be regarded merely as a very convenient way of tying together a large group of important physical phenomena, but not as being by any means a full or complete statement of natural relations.

With respect to the general theory, however, I believe the situation is quite different. The fundamental postulate that the laws of nature are of invariant form in all coördinate systems is highly mathematical, and of an entirely man-made character. Of what concern of nature's is it how man may choose to describe her phenomena, and how can we expect the limitations of our descriptive process to limit the thing described? Furthermore, Einstein's method of connecting his mathematical formulation and nature by way of coincidences of 4-events (three space, one time coördinates) seems to be very far removed from reality, since it entirely leaves out the descriptive background in terms only of which the 4-event takes on physical significance. Nevertheless, three definite conclusions about the physical universe have been taken out of the hat by the conjuror Einstein (shift of the perihelion of Mercury, displacement of apparent position of stars at the edge of the sun's disk,



and the shift toward the infra-red of spectrum lines from a source in a gravitational field), and the problem for us as physicists is to discover by what process these results were obtained.

An examination of what Einstein actually did in deriving his results will show, I believe, that the situation is really different from that suggested above. In the first place, the requirement that the laws of nature be of invariant form actually places no restriction, as any one can see by setting himself the task of expressing, for example, an inverse cube law for gravitation in terms of generalized coördinates. The work of expressing such a law can be attacked in a perfectly routine way. (The essential difference between the invariability requirements of the special and general theories is to be noted; the special theory requires that the velocity of light, for instance, have the same *numerical* value in all allowed systems: the general theory merely that all laws have the same *literal* form, but with variable numerical coefficients.) But, as Einstein says, if any one actually attempts to carry through the work of expressing an inverse cube law in generalized coördinates, he will find the task prohibitively complicated, and will seek for some simpler formulation. What Einstein actually did, therefore, was to require that the laws of nature be *simple* in generalized form. Now we know that the law of gravitation as formerly expressed in ordinary coördinates as an inverse square law was approximately exact, and was also simple. Any deviations from this law are small, and all experience leads us to expect

that to the first order of small quantities the deviations can be taken care of mathematically in the form of small correction terms to this law. This by itself gives nothing, however, because a small correction term can be added to our equations in an infinite number of ways. If, however, we know that the equation must be of a certain type after the correction terms have been added, the possibilities are so much restricted that the form of the correction term may be determined. In arguing as to the probable type of the equation, Einstein advanced the considerations by which physics gets into the situation.

In the first place, the special theory had prepared us for the possibility of finding that our measuring instruments might be modified in a gravitational field, analogously to the shortening of a meter stick when set into motion. In fact, special theory had indicated that in an accelerated system the modifications might be too complicated to be treated by that theory. In the absence, then, of specific information we must be prepared for the most general possible alteration in space-time in a gravitational field. In describing space-time we must therefore use coördinates adapted to handling the most general possible relations, and these are the generalized coördinates of Riemann, which had been already discussed by mathematicians. Going back now to Einstein's criterion that the equations are to be simple, we have the demand that the equations be simple in generalized coördinates, and of course they must also reduce to the ordinary equations (that is, the equations of spe-

cial relativity) in space where there is no gravitational field. In deciding the further question as to what the type of equation probably is, we are influenced by considerations of convenience as well as by physical considerations. Practically the only type of equation that can be handled mathematically is linear, so that we shall certainly try first whether this type of equation may not continue to hold. Now the Newtonian law of the inverse square may be expressed in terms of a linear differential equation of the second order in the old Cartesian coördinates (Poisson's equation), so that our most immediate suggestion is that the equations remain linear and of the second order in generalized coördinates. As a matter of fact, this requirement turns out to be sufficient to determine the small correction term by which the ordinary equations can be generalized; Einstein's papers must be studied to see how this works out in detail.

All this looks pretty mathematical, but as a matter of fact there is much physical content, because systems which can be described by linear equations of the second order have definite physical properties. The requirement that the equations be linear corresponds to one of the most fundamental properties of our universe—the causality concept would not be possible or would be much modified in a universe governed by non-linear equations, for the joint effect of two causes acting together would not be the sum of their effects acting separately, so that the analysis of a situation into simple elements would be impossible

and the causality concept probably would not have arisen. Furthermore, an equation of the type of Poisson of the second order means that there are propagation phenomena, and equations of mechanics of the second order involve the existence of a scalar energy function. If, then, the behavior of the universe can be described by differential equations at all, these equations must be linear of the second order if the universe is to have the broadest physical characteristics of our own universe. What Einstein really did, therefore, was to demand that even when space-time is warped by the presence of a gravitational field, those physical phenomena which can be described in terms of differential equations continue to be described by linear differential equations of the second order; that is, that nature continues to be describable in terms of a causality concept, with propagation phenomena, and a simple energy function. The consequences of a guess like this about the properties of nature appeal to our physical intuition as being worth following out, and of course we know the experimental justification.

Several general comments may now be made on the structure reached in this way. In the first place, the whole structure is only descriptive in character; we find certain correlations in nature which we describe with considerable completeness in mathematical equations, without introducing any new element of explanation or of mechanism. We have seen that as we increase our range from the realm of ordinary phenomena to phenomena of different character we

arrive at a stage where for a time the process of explanation apparently halts, and we have to be satisfied with a statement of mere correlation between elements; later, however, these elements may be accepted as the ultimates in a broadened scheme of explanation, and the explanatory process resumed. Are we at such a stage now with the general relativity theory, and may later a new scheme of explanation be established based on the correlations of Einstein? This is of course a matter of individual judgment; I personally question whether the elements of Einstein's formulation, such as curvature of space-time, are closely enough connected with immediate physical experience ever to be accepted as an ultimate in a scheme of explanation, and I very much feel the need for a formulation in more intimate physical terms.

In the second place, we must repeat the comment already made in discussing time, namely, that there is still a very wide gap between the theory and its physical application, in that we have no way of identifying our physical clocks and our physical measures of time with the thing called time in the formulas. This gap must be filled by a specification of the physical structure of a clock.

It has always been very puzzling to understand why Einstein has so strenuously insisted that the shift toward the infra-red is an integral part of the general theory, and that if the shift is not found, the theory must fall. In other words, Einstein insists that the assumption that an atom is a clock is an integral part



of his theory. I believe that this attitude may be due to a realization by Einstein of that very flaw in the logical structure which we are now emphasizing. In the absence of any method of specifying the details of construction of at least one clock, relativity becomes a purely academic affair, unless there exist in nature concrete things which may serve as clocks. Einstein *must* either be able to tell how to construct a clock, or else be able to point to a specific example of a clock. He chose the atom as the specific thing. Doubtless the reason was the apparent simplicity of the vibrating mechanism of an atom, as shown by the precise equality of the frequencies emitted by all atoms of the same element. If the atom is not a clock, where in nature can one be found? But in the last few years we have come to appreciate the exceedingly complicated quantum structure of an atom, and Einstein's thesis loses much of its instinctive appeal.

Since Einstein created the theory of relativity, it is perhaps ungracious to question his right to stipulate that the assumption that the atom is a clock is an integral part of the theory. This, however, degenerates to a mere matter of language, and does not touch the arbitrary nature of the procedure. It does not prevent us from having a second brand of relativity theory, that of X instead of Einstein, exactly like that of Einstein except that perhaps now the "clock" is constructed in terms of the life period of a radioactively disintegrating element. The only way to eliminate the arbitrariness seems to be to postulate that *all* natural processes, which run naturally of

themselves independently of what we may do, may equally well serve as clocks and give the same results. But in answering the question of the operational meaning of "independently of what we may do" we shall effectively have to answer the question of what is a clock. This point of view may possibly, however, get us a little nearer to our goal of finding how to specify the structure of a clock.

Finally, the general theory is not completely general, but applies only to a certain range of phenomena, just as we saw that the special theory does not embrace all optical phenomena. The general theory applies only to those phenomena which can be described in terms of differential equations, that is, *par excellence*, to large scale phenomena. If quantum phenomena cannot be described by differential equations,<sup>1</sup> as apparently now they cannot, general relativity cannot by its very nature be applicable. General relativity does not give us a comprehensive formulation of the behavior of all nature, and as far as we can see, we are still as far as ever from such a general formulation.

### ROTATIONAL MOTION AND RELATIVITY

Physically there is a great difference between the behavior of systems in uniform relative rectilinear motion and those in uniform relative rotation. The special theory of relativity states that there is a triply infinite number of systems with all possible uniform

<sup>1</sup> This statement now takes on a very questionable aspect in view of the new quantum wave mechanics (March, 1927).

rectilinear velocities with respect to each other, in all of which physical phenomena have exactly the same mutual relations, that is, natural laws are the same. Now the mere formulation of the principle suggests the sense in which "system" is here used. It is obvious that "system" refers only to a part of the universe; we are not making a hopelessly academic statement about what would happen if we had an infinite number of universes to experiment with, but are talking about operations that may be approximately realized in our own universe. The "system" of the formulation we may think of as a completely equipped laboratory, out in empty space, so far from the heavenly bodies that they can have no effect. The different systems of the formulation are different laboratories, all built to exactly the same architectural blue prints. The phenomenon to which the postulates of relativity apply are phenomena which pertain entirely only to one or another of these laboratories. The meaning of this restriction is not completely definite and has, in any special case, to be judged partly by the context. Obviously, to see from the window of a laboratory another laboratory passing with a certain velocity cannot be counted as one of the allowed phenomena. Still less is it one of the allowed phenomena to observe that the center of gravity of the entire stellar universe has a certain velocity of translation with respect to the laboratory. The special principle of relativity contains by implication therefore the statement that certain very large and important classes of physical phenomena may be isolated

and treated as taking place unaffected by the rest of the universe. Granted now the possibility of isolation, we have a second statement, which is usually treated as if it were the entire statement of the restricted principle, namely, that there is a triply infinite set of systems in which these phenomena run in the same way independent of the relative motion of the systems with respect to each other. When once the significance of the observation is grasped that absolute motion has no meaning in terms of operations, we see that this last statement takes immediately a most simple and satisfying aspect, in fact, so simple and inevitable that we are inclined to see in this the complete essence of the situation and regard the meaninglessness of absolute motion as affording peremptory proof of the restricted principle.

With this bias we now turn to examine the facts of rotary motion, and are disconcerted to find them quite different. No meaning in terms of measuring operations can be given to absolute rotary motion any more than to absolute translation, but nevertheless phenomena are obviously entirely different in different systems in relative rotary motion (phenomena of rupture, for example), so that apparently there are physical phenomena by which the concept of absolute rotary motion might be given a certain physical significance. Given two worlds like our own in empty space, but surrounded by impenetrable clouds, and each provided with a Foucault pendulum, then we believe that it is physically possible that we may

find on one of these worlds the plane of rotation of the pendulum gradually changing in direction, while on the other it remains stationary. This difference we regard as possible without other accompanying physical phenomena which are causally related to the rotation of the pendulum (of course we have to make the two worlds of infinitely rigid material and eliminate other phenomena which we regard as purely incidental), so that we apparently have here a contradiction of our cardinal physical principle of essential connectivity. We are certainly not inclined to give up our principle, and we believe that as a physical fact, if the clouds could be evaporated, an observer in one world would find that he was rotating with respect to the system of the fixed stars, whereas the corresponding observer on the other world would find that he was stationary. Our principle of essential connectivity is therefore maintained, in that the rotation of the plane of the pendulum is connected with a rotation with respect to the rest of the universe of the entire world in which the pendulum is mounted. As far as I am aware, no other way of maintaining our principle has ever been suggested. But this demands that we give up our physical hypothesis of the possibility of isolating a system. There is here no question of limiting behavior; we believe that no matter how far our rotating world gets from the rest of the universe the Foucault pendulum would always behave in the same way; the system can never be isolated, but such local phenomena as the invariance of the plane of the pendu-



lum are always essentially determined by the rest of the universe.

If now our system cannot be isolated, we must return to the phenomena of translational motion. In principle the act of isolation cannot be performed, the rest of the universe cannot be disregarded, and we should expect that different states of translational motion as well as different states of rotational motion with respect to the rest of the universe would have an effect on phenomena. We set ourselves the problem of understanding this apparent enormous difference between phenomena of translation and rotation. We remark that what apparently is a difference in principle may, in virtue of the approximate character of all measurement, be only a difference in magnitude, and that translational effects may exist too small to detect. A physical basis for such a difference may be found in the enormously different numerical values of translational and rotational velocities with respect to the rest of the universe attainable in practice. In describing phenomena of cosmic magnitude, we may plausibly measure the phenomena in units commensurable with the scale of the phenomena. Thus in measuring linear distances, we may perhaps choose as the unit of length the diameter of the stellar universe, and in measuring rotation, a complete reversal of direction with respect to the entire universe. This last means a change of angular orientation of  $2\pi$ , the first means a length of the order of  $10^5$  light years. Measured in such cosmic units the angular velocities attainable in practice are

incomparably greater than linear velocities. We now see that it is possible that the real state of affairs is as follows: namely, phenomena in any system are affected by motion with respect to the entire universe, whether that motion is of translation or of rotation, and the magnitude of the effect is connected with the velocity of the motion by a factor which is of the general order of unity when velocity is measured in cosmic units. This last is merely an application of the argument so often made in physics as to the order of magnitude of unknown numerical factors, and will be found expanded on page 88 of my book on *Dimensional Analysis*. The linear velocities attainable in practice are now so exceedingly low that their effect has not yet been detected experimentally, but angular velocities are high, and the effect is easily demonstrable. In this light the special principle of relativity is no different in character from any other physical law; it is only approximate, and some day our measurements may become refined enough to detect its limitations.

We have made a hypothesis here, which we may call the hypothesis of the immanence of the entire universe, namely, that isolation is impossible, or that the rest of the universe, no matter how distant, always has a local effect on at least some phenomena. This is essentially the hypothesis of Mach,<sup>1</sup> and leads to a situation which can, I think, be contemplated

<sup>1</sup> E. Mach, *The Science of Mechanics*, translated by McCormack, The Open Court Publishing Co., Chicago, 1893. See especially p. 235.

with logical equanimity, although it has always seemed to many physicists most highly antiphysical in character. It must certainly be admitted that most physical experience justifies us in thinking that effects may be made as small as we please by getting far enough away from the cause of the effect. But if we accept the considerations of the preceding pages, we must be prepared to admit that as phenomena change in range their character may change, and that in these new realms we must, at first at least, be satisfied with a mere statement of correlations. Certainly we have very strong physical evidence of a formal correlation between the Foucault pendulum and the rest of the universe. But a correlation of this sort may be without significance because of its very breadth; we never can prove the significance of the correlation by performing an experiment with the rest of the universe absent. Have we really done anything more than merely get things into such a formal situation that they cannot be assailed, a possibility which the mere laws of our thinking seem always to leave open, as has been suggested, or is there any physical content to what we have done? We have seen that if our correlation is also suggested by other phenomena, then we may accept it as having physical content. Now there is just a glimmer of a suggestion that our hypothesis of the immanence of the universe may be needed in other ways. The gravitational constant and the velocity of light are always treated as arbitrary magnitudes thrust on the universe from outside

with no connection with other phenomena. Nevertheless, I suppose that no one regards this situation as ultimately satisfactory and does not entertain the hope that some day we may be able to give some sort of account of the numerical magnitude of these constants. We have not hitherto succeeded in finding any connection between these constants and small scale phenomena such as the charge on the electron, its mass, etc., so that there is some plausibility in expecting that a connection may be sometime found with cosmic things; indeed general relativity theory already prepares us for exactly this possibility. Now the velocity of light and the gravitational constant control small scale experiments, for of course these two constants can be measured by local experiments, so that if the cosmic connection is found, we should have a control of local behavior by cosmic things, and therefore another example of the immanence of the entire universe. There is no need for me to waste time in apologizing for the highly speculative character of all this. It is worth while to emphasize, however, that our general considerations on the meaning of "explanation" have prepared us to admit as reasonable just the sort of explanation contained in the hypothesis of the immanence of the universe, and therefore to reserve a place in our physical thinking for possibilities of this sort, in spite of the fact that such considerations are not usually entertained, and may seem to many opposed to the spirit of physics.

QUANTUM CONCEPTS<sup>1</sup>

The history of quantum theory up to the present is a repetition in many respects of that of the early theories of electricity, in that all our thinking has been in mechanical terms. As far as we now know, quantum phenomena are always associated with atoms. We make for the atom a mental model with all the properties of the mechanisms of the ordinary scale of magnitude and with a few impressed properties in addition which represent the new quantum relations. As we now think of it, the atom has a massive core about which electrons revolve under an inverse square law, the connection between the mass of the electron, its acceleration, and the force acting on it being that usual in Newtonian mechanics. The space in which the electron circulates is thought of as Euclidean, and the motion is described in time, which may be measured with clocks in the usual way. The general equations of electrodynamics do not apply; there are no propagation effects inside the atom, the motion of the electrons does not produce a magnetic field, and there is no radiation when the electron is in one of its possible stable states, in spite

<sup>1</sup> This section was written early in 1926 without access to recent literature. Our attitude toward quantum phenomena has been so much changed since then by the "new" quantum mechanics, that a number of the following statements are superseded as a statement of present opinion. However it has seemed worth while to let the section stand as written, because many of the developments actually taken in the new mechanics follow the lines that it is here urged they ought to take, and in so far afford interesting confirmation of the point of view of this essay.



of the acceleration. We may, if we please, in working out the character of the motion, entirely neglect the electrical origin of the inverse square law, and treat this merely as an impressed force without further implications. Superposed on the ordinary spatial, temporal, and mechanical characteristics of the model are additional quantum properties, one which determines the particular orbit in which the electron moves [ $\oint pdq = nh$ ], and another which determines the frequency of the radiation emitted when the electron passes from one allowed orbit to another. No mechanism is suggested to account for these quantum conditions, although the conditions are formulated in mechanical terms.

We now have to ask what is the meaning in terms of operations of our usual concepts of space-time and mechanics when applied to phenomena of this order. It is of course evident, as has already been emphasized, that the concepts have entirely changed in character, because we do not measure an electron orbit, for example, by stepping off the diameter with meter sticks, or by measuring the time required for light to travel across the diameter. The particular feature of immediate interest in this changed situation is the change in number of our concepts on the atomic level. I shall not attempt to find by an exact analysis the number of independent concepts at this level; probably such an analysis is not possible. We may, however, make an approximate suggestion. Apparently the most important concept in describing relations inside a quantum system corresponds to that

of energy on the ordinary scale. Changes of energy determine the frequency of emitted radiation, as well as the relations during collisions of atoms and electrons; these collisional relations make direct connection with experiment through the voltages applied to electrons in collision experiments. The analogue of the momentum concept also seems to have independent significance, as shown by the Compton effect. The frequency of emitted radiation is also something with independent experimental significance. I believe that these three things are all that have direct significance for quantum experiments made up to the present time. In any event, it is perfectly evident that on the quantum level the concepts which at present have operational significance are considerably fewer than on the level of ordinary experience.

Apart from the question of convenience, there may be justification in continuing to use our old mechanical forms of thought if new experimental relations are thereby suggested. That a very large number of such as yet undiscovered relations may be suggested in some such way is at once evident. Thus we have no present knowledge of any phenomenon associated with what the electron does when passing from one energy level to another. How long does it take to make the passage? What is its path during passage? Is it subject to the ordinary laws of electrodynamics during passage? When and where is the radiation emitted that corresponds to passage? When the electron leaves one stable orbit is the orbit on which it will eventually land already determined? Does the

radiation train emitted during a change from one energy level to another have a definite length in space, or may it have a variable length and correspondingly something that corresponds to variable amplitude? What happens to the radiation when the electron passages are interfered with before the emission of a quantum has been completed? What is the mechanism by which the quantum conditions are imposed? Is it not possible that part of the clue to the riddle of the manner of transition from purely quantum behavior to the behavior of classical mechanics may be found in the behavior of the electron during passage from one energy level to another? Certainly we have a tendency to the classical behavior under those conditions, such as at high temperature or in strongly condensed systems, in which the time occupied in passage might be expected to become a more important part of the total time.

Corresponding to these questions there should be many as yet undiscovered phenomena, and the mechanical point of view therefore has its value in suggesting experiments to detect such effects. It is of course too early to see what the final result will be here; we cannot tell whether eventually enough new experimental kinds of behavior will be found to restore the number of independent concepts to that of the level of ordinary experience or not, or whether indeed it will turn out that a greater number of concepts is required. It is contrary to our instincts to expect a greater number, and a smaller number now seems to us not unnatural, but the considera-

tions of this essay should prepare us for either possibility.

It is often said that quantum phenomena are inconsistent with ordinary mechanics, and proofs of this assertion are often offered. I believe that no such proof, in the spirit in which the attempt is usually made, can be correct, for it seems to me that the remark of Poincaré applies, namely, that any sort of behavior can be imitated by a mechanical system, provided it is only complicated enough. A peremptory proof of this can be given to any one who is not a believer in vitalism. If a sentient being can be regarded as a mechanical system, we merely have to station inside each atom a Maxwell demon, with *instructions* to make the atom react according to quantum rules. Opposed to the spirit of this sort of reduction of quantum phenomena to mechanical terms, we have to remember that it makes sense to talk about the character of our conceptual structure only when the number of concepts is reduced to the number that have independent operational significance, that is, to the minimum number.

In the meantime let us examine what may be the significance in the light of present experiment of statements like those ascribed to Bohr that our usual concepts of space and time may be inapplicable in dealing with quantum phenomena. This idea is often given the more explicit form that space and time may be essentially discontinuous at the quantum level. From the operational point of view, it is most difficult to see exactly what this more explicit state-

ment means, at least in terms of those operations by which length and time were originally defined. Thus if space were discontinuous, it might mean that a point exists which may be reached by laying off a meter stick fourteen times, for example, and another point by laying off sixteen times, but that no point can be found with fifteen applications. Such a state of affairs seems to be inconsistent with our definition of the counting operation and to have no concern with any properties of space; for what shall we mean by laying off a meter stick sixteen times if it cannot be laid off fifteen times? It is conceivable that space might end, in the sense that beyond a certain limit there might be some irremovable physical hindrance to the continued laying off of distances with a meter stick (although I think that we should be inclined to describe such a state of affairs in terms of matter enclosing empty space rather than as the end of space), but to say that space may be discontinuous seems to be meaningless. In the same way, I believe it meaningless to speak of discontinuous time. We may have phenomena discontinuous in space and time, but not discontinuous space or time.

It seems then that we must give up the idea that in the quantum domain the usual concepts of space and time may fail, in the specific sense that they may become discontinuous. What may we understand by the failure of these concepts in a more general sense? No one of course would expect that even eventually the concepts will have the same operational signifi-



cance for the inside of an atom that they have on the ordinary scale; it must be a modified sort of concept with which we are concerned, such as we have already seen is given by the field equations of electrodynamics. If now the number of operationally independent concepts on the quantum level turns out to be the same as on the level of ordinary experience, and if there is also the possibility of continuous transition from the operations of the quantum domain to those of ordinary experience, then it seems to me that we should say that our usual concepts of space and time still apply in the quantum domain. But if the number of operationally independent concepts is either greater or less than on the ordinary level, then I believe we must say that the ordinary concepts of space and time cannot apply. One might still look for the possibility of separating out from the complex of concepts on the quantum level a group which might change continuously to those of space and time on the ordinary level, but I think that such a possibility is very remote when one considers that the total number of concepts changes, and that in the zone where the number changes the definitions are not unique by which one extrapolates a concept from one domain to another.

If Bohr's idea is true that space and time cannot be used in describing ultimate quantum phenomena, one of the most immediate implications in terms of experiment might be that phenomena corresponding to intermediate positions of the electron between stable orbits do not exist.

Finally, we must comment on the general tactics of the quantum situation. It would seem that there have already been a sufficient number of unsuccessful attempts to formulate quantum behavior in terms of ordinary mechanics to justify the expectation that ultimately something quite different must evolve. The difficulties of an unmodified carrying over of ordinary mechanical notions to quantum phenomena may be illustrated by a simple example. Consider a particle of mass  $m$  rotating in a frictionless circular track of radius  $r$ . Then according to quantum conditions it can move stably on this track only with certain definite velocities, such that  $\int p dq = mv 2\pi r = nh$ . Suppose now the particle rotating with one of the allowed velocities, and a tangential force applied. If the usual mechanical notions of force are still valid, the particle must respond by moving in its track with continually increasing velocity. After the velocity has been increased by a small amount, we remove the force. The motion is now no longer one of the allowed ones, and the particle must in some way change its velocity; it must either slow down or speed up. In the first case it must either radiate energy, which a system of the simple mechanical properties we have supposed is not capable of doing, or else the law of conservation of energy fails, and also Newton's first law of motion during the process of acquiring the steady condition. If, on the other hand, the particle speeds up, it must increase its energy from nowhere, and again ordinary mechanics does not apply.

It seems then a mistake to attempt to formulate the quantum conditions in terms of the notions of ordinary mechanics (momentum, and position coördinates in either the ordinary or the generalized Lagrangean sense). It would seem, on the other hand, plausible to expect that mechanics is not a fundamental thing, but is in some way an effect produced by the aggregate action of a great many elementary quantum processes. Amplitude of radiational vibration, for example, may be such a statistical aspect of a great many processes, in some such way as on the ordinary level of experience temperature is a statistical aspect of the average kinetic energy of the atoms. One possibility of this kind has already been more explicitly indicated; in the elementary process of emission of radiation, frequency and energy are not two independently assignable variables, but are connected [ $E=h\nu$ ]. That is, on the quantum level radiation has only a single property, which is properly neither energy or frequency. [We are now neglecting the polarization aspect of radiation.] On a higher level, that of ordinary radiation, the single elementary property has expanded itself into two (energy and frequency) through the additional variable of the number of elementary quantum processes in the complex radiation.

The program of the immediate future should be an extension of something of this sort, namely, to invent new concepts corresponding to the experimentally independent things on the quantum level (such perhaps as the resultant of the fusion of the

energy and frequency concepts for radiation), and then to show how the ordinary concepts of mechanics (and very likely those also of space and time) are generated by statistical effects in aggregates of great numbers. Perhaps it is yet too early for an attempt of this sort, because it may seem that there are still too many possibilities of new experimental discoveries which might upset the results of elaborate theoretical speculation. If this should really be felt to be the case, I believe that physics ought for the present to hold in partial abeyance its theoretical activities in this field, and devote itself to acquiring as rapidly as possible the necessary experimental facts. We may emphasize again that the possibility of carrying out this plausible program can be proved only by experiment; it may be that more concepts will be required on the quantum level than for ordinary experience.

The invention of new concepts is certainly not an easy thing, and is something which physics has always deliberately, and perhaps justifiably, shirked, as shown by the persistent attempts to carry the notions of mechanics down into the finest structure of things. This shirking has not had bad results, but on the contrary good results, as long as physics has been primarily concerned with phenomena near the range of ordinary experience, but I believe that as we get farther and farther away from ordinary experience, the invention of new concepts will become an increasing necessity.

## CHAPTER IV

### SPECIAL VIEWS OF NATURE

IN this last chapter we propose to discuss certain special hypotheses about the structure of nature, and certain other matters that could best be left until we had examined our fundamental concepts.

We have seen that in setting up the general rules which are to guide us in describing and correlating nature, we have to take extreme care to allow no special hypotheses to creep in, as otherwise we might be restraining possible future experience. Even here there is no hard and fast line of separation of the general from the special, and one might entangle himself in inextricable difficulties if his ideals were too meticulous. How for example is the critic to be answered who says: "Your very endeavor to formulate principles so broad as not to restrict future experience means, when examined in the light of operations, that you are seeking for principles which *past* experience suggests will not limit the future. It is in the very nature of things impossible to escape all the implications of past experience and therefore to find any completely general principle." I believe that we must admit the critic is right, and that rigor-



ously our goal is impossible of attainment. We may say in partial self-defense that all the discussion of this essay has been subject to one explicit assumption, namely, that the working of our minds is understood, which of course involves the assumption that our minds continue to function in the future in the same way as in the past. Even with this proviso we can not rigorously avoid the implications of the past, but there can be no practical question that we recognize certain assumptions about the behavior of nature to be so special as to limit seriously the physical possibilities, and other assumptions to be less restricting. In the previous discussion we had to make assumptions, but I hope these assumptions will be recognized by all with physical experience to be so broad as not to restrict us seriously. More special assumptions or hypotheses have their very great use, however, when we attempt to push forward the domains of experimental knowledge, because they may suggest new experiments or aid in correlating information already obtained. These special hypotheses may cover a very wide range of generality; some of them are general enough in character to be discussed here.

Among these special hypotheses there is a group which play an important part in the speculation of most physicists, and which have features in common. These are: the hypotheses of the simplicity of nature, of the finiteness of nature in the direction of the very small, and of the determinateness of the future in terms of the present. That these views have points of similarity is obvious if we consider a hypo-

thetical special case. Suppose that no physical structure beyond the electrons and protons can be discovered, or is even suggested by any known phenomenon, so that the entire future behavior of a system can be determined by a specification of the present relations of all its protons and electrons; in this case nature would be both simple and finite and the future determined by the present.

### THE SIMPLICITY OF NATURE

Of these hypotheses, perhaps the most important is that of the simplicity of nature, because of its wide spread diffusion and the effect it has had on physical thought. The hypothesis of simplicity assumes several forms; some physicists are convinced that the laws which govern nature are simple, others that the ultimate stuff of which nature is composed is simple (perhaps protons and electrons and energy), or there may be a combination of both views into the belief that ultimately we shall find simple ultimate elements behaving according to simple laws. In one respect it is obvious that nature is not simple, namely numerically—try counting the electrons or atoms or stars!

Consider now the first of these aspects of the thesis of simplicity, which may be expressed as the conviction that the behavior of the entire universe can be comprehended in a few principles of great breadth and simplicity, such as the inverse square law of force, or the second law of thermodynamics, or perhaps still better the equality of the elementary posi-

tive and negative charges, which apparently holds to an enormous degree of precision. In explanation of a view like this there is in the first place the mental urge, because we can take a satisfaction almost æsthetic in contemplating such a universe, and there is in the second place a strong suggestion from experience. Practically all the history of physics is a history of the reduction of the complicated to the simpler. For example, the behavior of a large part of the world of immediate experience can be reduced to the simple laws of mechanics. The behavior of another very large group of natural phenomena can be reduced to thermodynamics. The behavior of the heavenly bodies, which at first was described in a rather complicated way in the Ptolemaic system of astronomy, can be reduced to those same laws of mechanics which we find in our immediate neighborhood, with the one addition of the universal law of gravitation, which later refined experiment discloses is really active in our immediate surroundings. Similarly the laws of thermodynamics (except that part dealing with radiation) are reduced to the ordinary laws of mechanics through the additional assumption of the atomic structure of matter. Truly a stupendous accomplishment that may well color our whole future outlook. One may find great justification here for the belief that all nature will ultimately be reduced to a similar simplicity, and, in particular, justification for the attempt to find the explanation of all nature in the action of mechanical laws. Now, of course, as a matter of physical and historical fact,

this program could not be carried through, but obdurate physical phenomena were discovered. Electric phenomena, which at first seemed so promising, refused to fit into the scheme, and the converse attempt, to explain mechanical effects in terms of electrical effects, also failed. We still carry our ordinary mechanical notions down into the realm of small electric effects, and still talk, for instance, about non-electrical *forces* which hold an electron together. Nor are there experiments affording sufficient basis for believing that all the mass of a positive nucleus is electrical in character. We also think of electrical charges as having the property of identifiability, which involves the possession of sharp edges and a change in the law of force at small distances, and this is certainly a property carried over from our large scale experience.

It seems fairly evident then that the laws of nature cannot be reduced to either those of mechanics or of electricity, nor probably, as is suggested by quantum phenomena, to a combination of both. This of course does not preclude the possibility that the laws still may be simple when expressed in other forms. An example of such a broad general law that goes deeper than mechanics or electrodynamics is probably afforded by the second law of thermodynamics when extended to include radiation phenomena. Examples of attempts to find other such simple laws are Tolman's Principle of Similitude,<sup>1</sup> and Lewis's theory of

<sup>1</sup> R. C. Tolman, Phys. Rev. 3, 244-255, 1914; 6, 219-233, 1915; 15, 521, 1920; Jour. Amer. Chem. Soc., 43, 866-875, 1921.

Ultimate Rational Units,<sup>1</sup> and his recently enunciated principle of Complete Reversibility.<sup>2</sup> The first two of these attempts I do not believe are successful, for reasons I have stated elsewhere,<sup>3</sup> the third also seems somewhat doubtful.

With regard to the general question of simple laws, there are at least two attitudes; one is that there are probably simple general laws still undiscovered, the other is that nature has a predilection for simple laws. I do not see how there can be any quarrel with the first of these attitudes. Let us examine the second. We have in the first place to notice that "simple" means simple to us, when stated in terms of our concepts. This is in itself sufficient to raise a presumption against this general attitude. It is evident that our thinking must follow those lines imposed by the nature of our thinking mechanism: does it seem likely that all nature accepts these same limitations? If this were the case, our conceptions ought to stand in certain simple and definite relations to nature. Now if our discussion has brought out any one thing, it is that our concepts are not well defined things, but they are hazy and do not fit nature exactly, and many of them fit even approximately only within restricted range. The task of finding concepts which shall ade-

<sup>1</sup> G. N. Lewis. Vol. 15, 1921 of the Contributions from the Jefferson Physical Laboratory, dedicated to Professor Hall, Cambridge, Mass.; *Phil. Mag.*, 49, 739-750, 1925.

<sup>2</sup> G. N. Lewis, *Proc. Nat. Acad. Sci.*, 11, 179-183, 422-428, 1925.

<sup>3</sup> P. W. Bridgman, *Phys. Rev.* 8, 423-431, 1916; "Dimensional Analysis," p. 105.



quately describe nature and at the same time be easily handled by us, that is, be simple, is the most important and difficult of physics, and we never achieve more than approximate and temporary success. Consider the example of time. The original concept of local time, which for long seemed satisfactory, turns out to be inadequate, and has to be replaced by extended time, which is so complicated that it is questionable whether we shall ever be able to grasp it with the confidence that we must demand in a useful concept (by "grasp" I mean intuitive command of all the implications of the operations which are involved). The concept has not yet been found which describes simply the temporal relations of the universe.

Not only are concepts hazy around the edges and so incapable of fitting nature exactly, but there is always the chance that there are concepts other than those which we have adopted which would fit our present phenomena. Finding concepts to fit nature is much like solving a cross-word puzzle. In the puzzle there may be some parts of the pattern which we fill completely and easily, but sometimes we find parts in which we can fill in everything except one or two obstinate definitions, so that we are sure we are on the right track, and rack our brains for the missing words, when with a flash of inspiration we see that the obstinate words can be fitted in by a complete change in those which we had already accepted. It may be that we are soon to witness a similar change in our concept of the nature of light. An important

difference between the cross-word puzzle and nature is that we can never tell when we have filled in all the squares in any of the parts of nature's puzzle; there is always the possibility of new phenomena which our present scheme does not touch.

Considering, then, the nature of our conceptual material, it seems to me that the overwhelming presumption is against the laws of nature having any predisposition to simplicity as formulated in terms of our concepts (which is of course all that simplicity means), and the wonder is that there are apparently so many simple laws. There is this observation to be made about all the simple laws of nature that have hitherto been formulated; they apply only over a certain range. We have not extended the laws of gravitation to small bodies, nor have we found that our electrical laws will work on a cosmic scale. It does not seem so very surprising that over a limited domain, in which the most important phenomena are of a restricted type, the conduct of nature should follow comparatively simple rules.

A tempting question is whether there may not be some laws of nature that are *really* simple, without relation to our mode of formulation, such as the law of the inverse square. I leave it to the reader to decide whether this question has meaning. In this connection it is possibly significant that the average physicist is strangely reluctant to tamper with the inverse square law. I find in myself a lack of sympathy, which I cannot justify by any of the consid-

erations of this essay, with attempts like the recent one of Swann,<sup>1</sup> for example, to explain a wide variety of hitherto obstinate effects by the assumption of slightly unequal departures from the inverse square law by the electrons and protons. Of course I hope that this feeling will turn out not to be prejudice, but will perhaps be justified by some such general observation as that a departure from the inverse square law so slight as by definition to be forever beyond detection by direct experiment is meaningless; but of this I am not at all sure.

We are now ready to consider the second respect in which nature may be simple, namely, because the material of which it is built may reduce to a few sorts of elements. In this discussion it will be convenient to consider also at the same time the more inclusive simplicity arising from simple laws acting on simple elements. The immediate question for us here is one of fact: does nature seem to be getting intrinsically simpler as we get toward small scale phenomena? There is much room for difference of opinion here; personally I feel that this expected simplicity is not in evidence, at least to the extent that we could desire. For instance, the fact that the electrons must have both electrical and mechanical properties is a straw in this direction.

It must also be remembered that a certain simulation of simplicity is inevitable as we approach the limits of experimental knowledge, whatever the actual structure of nature, for the mere reason that

<sup>1</sup> W. F. G. Swann, *Phys. Rev.*, 25, 253, 1925.

near the limit our possible experimental operations become fewer in number, and our concepts fewer also. The question which we are trying to answer has, therefore, its real meaning only in terms of the possible future. Do we believe that if we drive in our stakes at a certain point on our present frontiers, this point will gradually, as physics advances, become possessed of a continually richer experience, so that nature at this point will appear increasingly complicated? Or do we expect a termination of this process of expansion fairly soon? It seems to me that as a matter of experimental fact there is no doubt that the universe at any definite level is on the average becoming increasingly complicated, and that the region of apparent simplicity continually recedes. This, however, is not the opinion of all observers. Thus Bertrand Russell, in "What I Believe", page 10, writes, "Physical Science is then approaching the stage where it will be complete, and therefore uninteresting."

This is perhaps a particularly favorable epoch in the history of physics to urge the essential complexity of nature, because all our new quantum phenomena indicate a vast wealth of hitherto unsuspected relations on the very edge of the attainable. There is one aspect of quantum relations, as also of our ideas of the nature of the structure of the nucleus of the atom, which is particularly significant in this respect, namely, that we have to describe phenomena by statistical methods. Now a statistical method is used either to conceal a vast amount of

actual ignorance, or else to smooth out the details of a vast amount of actual physical complication, most of which is unessential for our purposes. There can be no doubt of the amount of ignorance that the statistical method conceals when applied to these phenomena, but there are also strong indications, particularly when applied to the nucleus, that it covers a vast amount of actual physical complications. The nucleus of a radium atom becomes unstable on the average every  $10^4$  years, which may be plausibly taken to indicate that every  $10^4$  years the radium nucleus gets itself into some particular configuration. Considering the time scale on which we suppose events in the atom to take place, and also considering the fact that radioactive disintegration seems unaffected by outside agencies, this would indicate a perfectly appalling amount of structure. We are similarly driven to statistical methods in quantum theory, as for example, in Einstein's analysis of the details of equilibrium between emitting and absorbing atoms and radiation.

In general, we cannot admit for a minute that a statistical method, unless used to smooth out irrelevant details, can ever mark more than a temporary stage in our progress, because the assumption of events taking place according to pure chance constitutes the complete negation of our fundamental assumption of connectivity; such statistical methods always indicate the presence of physical complications which it must be our aim to disentangle eventually.



It appears then that present experimental evidence makes very probable structures beyond the electron and the quantum; we may go even further and say that there is no experimental evidence that the sequence of phenomena in nature as we go to ever smaller scales is a terminated sequence, or that a drop of water is not in itself essentially infinite. (This statement contains by implications the meaning that we attach to infinite.) All the more, then, there is no evidence that nature reduces to *simplicity* as we burrow down into the small scale.

Whatever may be one's opinion as to the simplicity of either the laws or the material structure of nature, there can be no question that the possessors of some such conviction have a real advantage in the race for physical discovery. Doubtless there are many simple connections still to be discovered, and he who has a strong conviction of the existence of these connections is much more likely to find them than he who is not at all sure they are there, and is merely hunting for anything that may turn up. It is largely a matter of psychology. Everyone knows that the mere suggestion that a problem has a solution, or the knowledge that someone has already solved it, is often sufficient to suggest a relation that otherwise might not have been noticed. The chances are, therefore, that the relations between phenomena will be found by those who are previously convinced that the relations exist. The observation that most of the discoveries are made by men with particular sorts of conviction naturally strengthens the belief that their convictions are true.

But this picture has an obverse side. The man who is convinced that there is a relation where none exists may waste all his time in vain seeking for it. Granted that nature has no particular predisposition to simple relations, the conviction that there are such relations is, from the point of view of any one individual, as likely to be a hindrance as a help. From the point of view of physical society, on the other hand, it is desirable that there be such convictions, for in such a society there will be more discoveries than in a society without such convictions. We have here again the old conflict between the individual and society. As in all other similar conflicts, society will not be able to demand permanently from the individual the acceptance of any conviction or creed which is not true, no matter what the gain in other ways to society. If nature is not simple, physicists will not continue to believe that it is, even if such a conviction does increase the total number of discoveries. It is an impossible attitude to expect that one can maintain. Does this then mean that physics is to face a drab future, becoming continually more prosaic, with new discoveries ever rarer, made by a continually decreasing number of misguided but fortunate enthusiasts? There may be such a danger, but the greatest part of the danger is avoided if its nature is clearly recognized. One of the problems of the future is the self-conscious development of a more powerful technique for the discovery of new relations without the necessity for preconceived opinions on the part of the observer.

There is an aspect here of our physical research that is often lost sight of, namely, the small proportion of successful discoveries compared with the number of investigators. Certainly the number of unsuccessful attempts, even in the case of those fortunate individuals who make the great discoveries, is very much greater than the number of their successful attempts. (Faraday's reputed satisfaction with a 1/10% return comes to mind.) This must always be taken into account in estimating the probable chances of correctness of any new theory. With so many physicists working to devise new theories, the chances are high that many false theories will be found, in which a number of phenomena may apparently fit together into a new relation, but which eventually prove to be inconsistent with other phenomena, so that the proposed theory has to be abandoned. As physics advances and the number of investigators and the amount of physical material increases, one has to be more and more exacting in one's requirements of a new theory. One must be particularly on guard against numerical coincidences. An interesting chapter might be written on numerical relations which have been hopefully published, but later had to be abandoned as without significance.

### DETERMINISM

If we are right in supposing that physical evidence gives no warrant for the idea that nature is finite downward, we have not only repudiated the thesis of simplicity but we have also made a very important

observation on the other general thesis mentioned at the beginning of this chapter, namely, the thesis of physical determinism. By determinism we understand the belief that the future of the whole universe, or of an isolated part of it, is determined in terms of a complete description of its present condition. [What we mean by present condition will be discussed later.] It is popularly assumed that every physicist subscribes to some such thesis as this. But now if there is infinite structure even in a small isolated part of the universe, a complete description of it is impossible, and the doctrine as stated must be abandoned. It seems to me that all present physical evidence prepares us to admit this possibility. I suppose, however, that most physicists would subscribe to some modification of the original thesis, perhaps along the following lines. Given a description of an isolated part of the physical universe in the most complete terms that have physical meaning, that is, down to the smallest elements of which our physical operations give us cognizance, then the future history of the system is determined within a certain penumbra of uncertainty, this penumbra growing broader as we penetrate to finer details of the structure of the system or as time goes on, until eventually all but certain very general properties of the original system, such as its total energy, are forever lost in the haze, and we have a system which was unpredictable. I suppose that it is a further conviction of at least many physicists that by sufficiently refining our measurements, the amount of haze at any fixed point in

the future may be made indefinitely small, and many might even go further and hope by studying the haze (perhaps statistically) to obtain some inferential evidence of structure beyond that yet experienced. In fact it may be that this last contains the germs of the ultimate method of investigation, if we ever reach a stage when we can no longer refine our methods of measurement.

Determinism to the physicist is simply a way of stating certain implications of his conviction of the connectivity of nature. We have seen that the broadest possible statement of the thesis of connectivity is: Given two isolated systems with identical past histories up to a certain epoch, then the future histories will also be identical. The thesis of the determinism of the future by the present constitutes a specialization of this general thesis in that we suppose that identity of *all* past history is not necessary for identity of future behavior, but only identity of present condition. The general and the special thesis are not equivalent by any means: if past histories are identical then present conditions are also identical, but the converse does not necessarily hold at all.

Now I believe that the general thesis (which I suppose all physicists will admit, but whose truth is nevertheless subject to the verification of experience) gets turned into the special thesis by a feeling of somewhat metaphysical content, which we may perhaps state by saying that we can see no way by which the past can affect the future except through the present. We do not like to think of the effect of a



cause distant in the past jumping over the present and affecting the future without touching the present at all. It is the analogue of that attitude of mind to which action at a distance in space is inconceivable; just as it is difficult to conceive of a body here affecting a body there without in some way an action propagated through intermediate space, so we do not like to think of a past cause jumping over time and producing a future effect without some sort of continuity in the causal chain through all intermediate time.

So far our discussion has been purposely loose: it is evident that what we mean by "present state" is crying for definition. What is meant by this may depend somewhat on the specific hypothesis that one adopts about the structure of nature. Historically the conviction of future determinism has been most intimately associated with a mechanical picture of the structure of the universe, so that it may be well to begin from this point of view. Suppose the simplest possible system composed of point masses without structure, as in the kinetic theory of gases. What sort of specifications do we believe necessary to fix the present state of such a system? The mechanical view of nature gives a definite answer. By present state we mean the positions and velocities of all the masses. This is sufficient for the complete determination of any purely mechanical system, in which the forces between the elements are known functions of only their relative positions. By a sort of extension of these ideas valid for mechanical sys-

tems, it seems to be often thought that the present state of *any* system is determined by a complete specification of the positions and velocities of all the ultimate elements of the system (provided always of course that this number is finite). This principle, however, does not appear to bear the check of experiment when applied to electrical systems with radiation. The theorems of the retarded potential show that such systems are determined by the present position and velocities of the charges in the immediate vicinity, and by the corresponding data at remote points given for proper epochs in the past; in this case, therefore, past and present history are necessary to determine the future. But if we consider the electrical field as part of the system, we may fix the future in terms of the present positions of the charges, their velocities, and the values of the field vectors all over space, thus returning to a certain formal resemblance to mechanical systems, and suggesting a reason for ascribing physical reality to the electric field. This analogy with a mechanical system is, however, loose; complete analogy would allow the instantaneous values of the time derivatives of the field to be given also, and this is not possible.

How is it that velocity can strictly be regarded as characteristic of the *present* state of the system? Certainly the usual operations for measuring velocity demand that we know the configuration of the system at two different times, and calculate the velocity from certain differences of the system at these two times. The velocity is defined as a limiting result, but even

in the limit the essential physical fact does not disappear that we must know the positions of the system at two times. We may now go further; if the velocity is properly included in the present attributes of the system, we can see no reason for not including a specification of all the higher time derivatives also. In the case of the simple gaseous system under present consideration we can answer this question by examining the operations by which we actually go to work to determine the future of such a system. The problem of determining the future condition of such a system reduces to the problem of writing the differential equations of motion of all its parts. If the system is a mechanical system, as in this case, these equations are of the second order in the time derivatives of the position coördinates, and also involve the forces, which we suppose are known in terms of the relative positions of the parts of the system. Given, then, the positions and the way in which the forces depend on the relative positions of the parts, the equations of motion can be written down for any configuration of the system, and these equations may be integrated (at least approximately) in terms of the proper initial conditions. Now the only boundary conditions on a second order equation are the initial positions and velocities. This is the reason that velocities have to be specified in giving the present condition of the system, and that it is not necessary to give the higher derivatives. Apparently the reason why we instinctively include velocity among the present properties of the system is not because

velocity is by its nature strictly a present property of the elements of the system, but rather because our wide experience with mechanical systems has shown that as a matter of fact velocity is necessary in such systems to determine future motion.

But now if the equations of motion of the parts of the system are not those of mechanics, they will in general be much more complicated in appearance and will involve higher derivatives of the time than the second. Suppose for the moment that the equations contain only derivatives and the mutual positions of the parts of the system. Then to integrate the equations and determine the motion we have to know the initial positions and the initial values of all the derivatives up to an order one lower than the highest which occurs in the equations. The equations of motion of an electron are even more complicated than this, in that the positions of distant parts of the system have to be given throughout an interval of time instead of merely an instant. It would seem that the feeling that the present state of a system may be determined in terms of positions and velocities does not as a matter of fact apply to all the systems of our experience.

The discussion up to this point has been subject to the fundamental assumption that the behavior of the system is entirely determined if we can give the position of each part as a function of time. This assumption is implicitly contained in Einstein's formulation of the general principle of relativity, namely, that there is nothing more to a physical system than a set

of space-time coincidences, and that the system is fixed in terms of the space time coördinates of all its parts. Already in discussing the assumption of relativity we have indicated reasons for dissatisfaction with this as a means of reproducing all experience, because in giving only the space-time coördinates of events we have entirely omitted the descriptive background of the equations, which gives physical color to the system in question. This discussion also assumes that a specification of the positions, velocities, and higher derivatives (if necessary) of the elements of the system is possible, which amounts essentially to the assumption that the system contains only a finite number of elements. Now in view of the experimental fact that there is no reason for supposing that the structure of the universe is finite, this conclusion must be modified, but I do not believe that the necessary modification affects the essential argument. In view of the possible infinite structure it would seem that we cannot expect more than that the future is determined by the present within a certain penumbra of uncertainty, and this penumbra may be made less important by digging down deeper into the structure when specifying the present condition.

We have also slurred over the ambiguities in "present" condition when the system is spread over space. Probably a unique ascription of meaning to "present" is not possible for an extended system, but at least one possibility is indicated by relativity theory. Imagine a staff of assistants distributed throughout space, each equipped with clocks syn-



chronized and set with the master clock by light signals in the conventional manner, and each fully equipped with the necessary measuring instruments. Then what we mean at this point of the argument by "present" state of the system is the aggregate of all the information about the *positions* and *velocities* of the ultimate elements which I determine in my immediate vicinity at my origin of time plus the reports of similar observations made by all the assistants, each local observation being made at the time origin of each local clock.

Going back now to the main argument, we have shown that the feeling that the present condition of the universe may be specified in terms of positions and velocities arose from experience with purely mechanical systems, and that the more general formulation, in which we add to the velocities the higher time derivatives, applies only to systems in which the ultimate elements move according to differential equations of higher order than the second. Furthermore, our analysis seems to have shown that systems in which there is radiation do not allow a determination of the future in terms of a present condition specified in terms such as these. It seems, however, that the general principle of the determinism of the future by the present may be saved by a change in the definition of what we mean by the present condition of the system, ridding it of its mechanical and other special implications, and making more immediate connection with direct experiment. Let us understand by present condition of a system the

aggregate of all information that can be obtained by any physical means whatever, with any sort of physical instrument, not attempting to get out of this analysis information about hypothetical ultimate physical elements, with the proviso that the measurements are to be made *now*, extending the concept of "now" to points distant in space in the way intimated above. With such a general definition of the meaning of "present" we can now deal with systems in which there is radiation, noticing that our assistant observers must be stationed throughout apparently empty space as well as in the neighborhood of matter. That this does adequately cover the case of radiation is suggested by considering again the two systems of dark lanterns with screens and distant mirrors which we have previously considered, in one system a light signal having been despatched 0.5 second ago and in the other 1.5 seconds ago. Our thesis demands that there be some present difference in these two systems, because their future history is different, in one of them a light signal arriving after the lapse of 1.5 seconds, and in the other after only 0.5 second. Now there *is* a present difference as reported by our assistants, for the assistant stationed half way between lantern and mirror reports in one system a flash of light on the side of a screen which is turned toward the lantern, and in the other system on the side of the screen turned toward the mirror.

This more general point of view answers the question whether velocity may be regarded as a present

attribute of the system, for the parts of a system which are in motion have momentum, and momentum may be detected by placing against such parts comparatively rigid members which will receive a minute deformation, so that velocity has a meaning in terms of physical measurements made at a single instant of time.

There is a subtle and difficult question here, namely, whether in talking about operations of measurement we can ever get rid of temporal implications, and therefore, whether a condition of the system in which temporal implications remain can properly be described as "present." I shall not attempt to answer this question: there must be some practically satisfying answer, involving perhaps the physical analogue of differentials of different orders in mathematics, short of carrying the analysis to such a degree of refinement that the concept of present becomes meaningless, as we can see might easily happen.

With this enlarged understanding of what we mean by present state of the system, it seems to me that physical evidence is now rather favorable to the view that the present determines the future, subject to qualification about the penumbra, at least as far as large scale phenomena are concerned. It appears much more doubtful when we come to small scale phenomena, and in particular it is doubtful whether the principle can be applied to the details of the quantum process, and in fact it is not certain that it has meaning. It is certain that if it is true an enor-

mous amount of structure beyond any that has yet been detected is implied.

### ON THE POSSIBILITY OF DESCRIBING NATURE COMPLETELY IN TERMS OF ANALYSIS

There is a certain thesis that is loosely related to the view that nature is finite downward, namely, that an explanation of the universe is possible in which we start with small scale things, and explain large scale phenomena in terms of their small scale constituents, the thesis, in other words, that all the properties of the large are contained in the properties of the small and that the large may be constructed out of the small. Some such thesis as this seems implied in the general attitude of many physicists. Let us examine the physical basis for this. To maintain this thesis would demand that aggregates of things never acquire properties in virtue of their numbers which they do not already possess as individuals. Is this true? Consider, for example, the two-dimensional geometry on the surface of a sphere. This is non-Euclidean. Is the geometry of the individual elements of the surface of the sphere non-Euclidean, or do they acquire this property in changing scale? Is the kinetic energy of a number of electrons all moving together in such a way as to constitute an electric current the sum of the kinetic energies of the individual electrons, or is there an additional term? Is the mass of an electron the sum of the masses of its elements?

A mathematical consideration is suggestive here. Those properties of a system which can be described in terms of linear differential equations have the property of additivity; the effect of a number of elements is the sum of the effects separately, and no new properties appear in the aggregate which were not present in the individual elements. But if there are combination terms (as in the electrical energy, which contains the square of the field), then the sum is more than (or different from) its parts, and new effects may appear in the aggregate. Now of course the linear equation is of enormous importance in describing nature, but many examples of systems with other types of equation can be found, as that above for electromagnetic mass. In expecting to find in nature such non-additive effects, we need not commit ourselves at all to the view that nature is governed by differential equations, but by analogy may expect similar effects if difference equations, for instance, should prove to be fundamental, or even something beyond present mathematical formulation.

It is certainly very much easier to handle a system physically if the total action can be built up from that of its parts, because the analysis which establishes the connection between the elements is easier to perform. It is obviously easier to show that an explanation in such terms is correct, because we have seen that explanation involves making experiments with representative elements absent or altered, and it is easier to vary the small things than the large things.



Those explanations which involve working from the small up will therefore be made first, and will appear to be of disproportionate importance. Places where I look for an explanation from the large to the small are perhaps in accounting for the values of the gravitational constant and the velocity of light and in those phenomena which general relativity theory indicates may depend on all the matter in the universe, as the Foucault pendulum experiment. We must, of course, also be prepared for such non-linear effects in the domain of unexplored quantum phenomena.

### A GLIMPSE AHEAD

Some of the general considerations of this essay may, with considerable plausibility, be expected to play a part in the future of both speculative and experimental physics. The most important effect may be expected from the clearer recognition of the operational character of our physical concepts. Indeed during the writing of this essay there has been a very marked increase in emphasis on the necessity of understanding in terms of physical operations such fundamental concepts as that of the electron, by the new quantum mechanics [the mechanics of Heisenberg-Born and Schrödinger of 1925-26].

We are to expect then in the first place a more self-conscious and detailed analysis of the operational structure of all our physical concepts. [It has been beyond the scope of this essay even to begin to attempt a systematic and thoroughgoing analysis of this

character.] This future analysis will show precisely how, as we extend the range of experience, the physical character of the operations changes by which we define our concepts, as, for example, in mechanics the notion of force disappears at high velocity and is replaced perhaps by the notion of momentum. In the region of change in the nature of our concepts, special study will be made of the accuracy of our physical measurements, and new experiments devised of greater accuracy, in order that we may know precisely to what extent the new concepts are equivalent to the old. Past experience suggests that we may perhaps expect to find new phenomena especially in those regions where the difficulty of carrying out the usual operation forces us to change the operational character of our concepts. There will be questions of a more or less formal nature to answer, as for example the *best* way of extending concepts when there are several possible courses open to us.

We may expect more interesting results, however, when we get so far beyond ordinary experience that the character of the possible physical operations has become so restricted as to result in an apparent decrease in the *number* of independent concepts. It seems plausible to expect that the structure of nature is more fundamentally connected with the number of independent concepts necessary for a complete description than with the precise details of the structure of the individual concepts, such, for example, as whether space is measured optically or tactually. In those regions where the number of

concepts decreases, we must make the most thorough-going experimental examination to discover if possible new sorts of operations by which the number of concepts may be brought back to normal. In searching for such new experimental operations it seems to me that by far the greatest promise for the immediate future is offered by improvements in our powers of dealing with individual atomic and electronic processes, such as we now have to a limited extent in the various spinthariscopes methods of counting radioactive disintegrations, or Wilson's  $\beta$ -track experiments.<sup>1</sup> In this self-conscious search for phenomena which increase the number of operationally independent concepts, we may expect to find a powerful systematic method directing the discovery of new and essentially important physical facts.

We can only conjecture whether the number of fundamental concepts will prove impossible of further increase or not, but present experience seems to give greater probability to the view that as we penetrate deeper the number of fundamental concepts will always tend to become fewer. We have already certainly one example in that the temperature concept disappears when we get to the atomic scale of magnitude, and possibly a second example in the building up of separate concepts for energy and frequency by the combination of great numbers of that one operationally simple thing which characterizes

<sup>1</sup> C. T. R. Wilson, Proc. Roy. Soc., 87, 277, 1912.

the elementary quantum process in ordinary radiation.

Different sorts of relations between concepts are conceivable in the transition zone where the number changes. We may find that other examples are like that of temperature, which is simply a statistical effect of a great many phenomena which may be described individually in terms of the ordinary concepts of mechanics, so that in this case the number of concepts changes merely by temperature dropping out, leaving the others more or less unaffected. Or all the concepts may be more closely interwoven, so that when the total number of concepts changes it may not be possible to separate out a group of concepts whose defining operations are unchanged. In such a case we must say that the original concepts are not applicable on the new level. The most immediate application of this idea has been already mentioned, namely, to the concepts of space and time. If the operations by which space and time are measured on the ordinary scale of magnitude cannot be carried down as a whole into the region of quantum phenomena, then we must say that the ordinary concepts of space and time are not applicable to these phenomena.

Closely connected with the sharper analysis of the operational structure of our concepts, we may expect in the future also a closer analysis of our inventions. This will take the form of a search for new physical facts which shall give to our inventions the character

of physical reality. In case prolonged search fails to disclose such phenomena (as is probably now the case with the field concept of electrodynamics), we must then find some way of embodying explicitly in our thinking the fact that we are dealing with pure inventions and not realities.



## INDEX

- Absolute, 26
- Absolute time, 4
- Action at a distance, 46
- Analysis of large into small, 51, 220
- Arithmetic, 35
- Atom, 59
  
- Bell, 84
- Birkhoff, 72
- Black body, 112
- Bohr, 190, 192
- Born, 222
- Boscovitch, 46
- Bothe and Geiger, 116
- Bridgman, 201
- Brownian movement, 107, 129, 143
- Bush, 142
  
- Caloric fluid, 59
- Campbell, 117
- Carnot engine, 125
- Causality, 80 ff
- Causal train of events, 85
- Clifford, 28
- Clock, 70 ff, 176
- Compton, 116, 188
- Continuity, 94
- Conservation of charge, 135, 136
- Conservative functions, 113
- Constructs, 53
- Correlation, 37
- Cosmic units, 182
- Cross word puzzle, 202
  
- Descriptive background, 64
- Determinism, 114, 209 ff
- Discontinuous space, 191
- Doppler effect, 166
- ds, 72
  
- Eddington, 93
- Einstein, VII, 1, 2, 3, 4, 7, 8, 9, 12, 13, 14, 64, 100, 155, 162, 163, 164, 167, 169, 170, 171, 172, 173, 175, 176, 177, 206
- Electrical concepts, 131 ff
- Electrical explanation of universe, 50
- Electrical mass, 139
- Electric field, 56, 133
- Empiricism, 3
- Energy, 108 ff, 126 ff
- Euclidean space, 14, 15, 16, 18, 23, 52, 61, 67
- Event, 95, 167
- Explanation, 37
- Explanatory crisis, 41
- Extended time, 77
  
- Faraday, 44, 57, 58, 209
- Final explanations, 48
- First law of thermodynamics, 126 ff
- Force, 102 ff
- Foucault pendulum, 180, 184
- Fourth dimension, 74
- Future, 222
  
- Gauss, 15, 34

- "Go and come" time, 112  
 Gravitational constant, 91  
  
 Haldane, 25  
 Heat flow, 130  
 Heisenberg, 222  
 Hertz, 44  
 Hoernlé, preface  
  
 Identity, 91 ff  
 Isolation, 82  
  
 Joule, 124  
  
 Kelvin, 45, 110  
 Kinetic theory of gases, 40  
  
 Lagrangean equations, 112  
 Larmor, 149  
 LaRosa, 164  
 Length, 9 ff  
 Lewis, 166, 201  
 Light, 150 ff  
 Local time, 75  
 Lorentz, 143, 147, 148, 149  
  
 Mach, 183  
 Mass, 102 ff  
 Mathematics, 60 ff  
 Maxwell, 44, 58, 112, 137, 148  
 Meaningless questions, 28 ff  
 Measurement approximate, 33  
 Mechanism, 45  
 Mercury, 105  
 Michelson, 15, 26  
 Michelson and Morley, 66  
 Models, 45, 52  
  
 Newton, 4, 110  
  
 Operational character of concepts, 5  
 Operational thinking, 32  
 Optical space, 67  
 Ostwald, 109  
  
 Penumbra, 34  
 Perrin, 107  
 Physical reality, 59  
 Planck, 69  
 Poincaré, 48, 115, 116, 190  
 Pythagoras, 61  
  
 Quantum act, 156  
 Quantum theory, 40, 47, 186  
  
 Radiation and temperature, 123  
 Relative character of knowledge, 25  
 Relativity, 150 ff  
 Reynolds, 93  
 Rotational motion, 178  
 Russell, 205  
  
 Schrödinger, 222  
 Space, 66  
 Silberstein, 11  
 Simplicity of nature, 198  
 Simultaneity, 7, 8  
 Spring balance, 103  
 Statistical methods, 115, 117  
 Stress, 54  
 Swann, 204  
  
 Table top, 106  
 Tactual space, 67  
 Temperature, 118 ff  
 Thermodynamics, 117  
 Thing traveling, 101, 152, 157, 164  
 Time, 69  
 Tolman, 200  
 Truth, 78  
 Turbulent motion, 120, 124  
  
 Velocity, 97 ff, 213 ff  
 Velocity of light, 100  
  
 Whitehead, 167  
 Wilson, 224













